

100

```

Qy 1741 atactaaactgggtgccttttgtaggaacacctgatttccttccatcctgagcgctctctctctcta 1799
|||||
Db 1741 atactaaactgggtgccttttgtaggaacacctgatttccttccatcctgagcgctctctctcta 1799
|||||

RESULT      2
AAK91990
ID   AAK91990 standard; DNA; 1230025 bp.
XX
AC   AAX91990;
XX
DT   13-SEP-1999 (first entry)
XX
DE   Nucleotide sequence of the complete genome of Chlamydia pneumoniae.
XX
KW   Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW   sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW   vaccine; neutralising epitope; ss.
XX
OS   Chlamydia pneumoniae.
XX
PN   WN0927105-A2.
XX
PD   03-JUN-1999.
XX
PF   20-NOV-1998;    98WO-I501890.
XX
PR   04-NOV-1998;    98US-O107078.
XX
PP   21-NOV-1997;    97FR-0014673.
XX
PA   (GEST ) GENSET.
XX
PI   Griffais R;
XX
DR   WPI; 1999-357842/30..
XX
PT   Genome sequence of Chlamydia pneumoniae
XX
PS   Claim 1; Page 291-611; 1912pp; English.
XX

The present sequence represents the complete genome of Chlamydia
pneumoniae, and encodes proteins AA34584-Y35879. C. pneumoniae causes
respiratory disease such as pneumonia and bronchitis and is thought
to be a contributing factor in heart disease, sarcoidosis, sinusitis,
purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
encoded by the open reading frames of the C. pneumoniae genome (See
AA34584-Y35879) can be used in immunogenic compositions as vaccines.
Vectors containing C. pneumoniae nucleotides sequences can also be
used as immunogenic compositions, especially where the vector directs
the expression of a neutralising epitope of C. pneumoniae.

SQ Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

Query Match          99.3%; Score 1786.4; DB 20; Length 1230025;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1798; Conservative      0; Mismatches     1; Indels     1; Gaps     1

QY 1 acttccccctgcataactatgctcagataatgctgatgattgcaggctcaggaggag 60
DB 244380 acttccccctgcataactatgctcagataatgctgatgattgcaggctcaggaggag 60
|||||
QY 61 aaattttcaaaaaaacctagtatttcoggaaattcgatatgcgcagaatatcatgtgg 120
|||
DB 244440 aaattttcaaaaaaacctagtatttcoggaaattcgatatgcgcagaatatcatgtgg 120
|||||
QY 121aatctgtataaccattctcttaagcctccgtagtcctccaagctgcaaggagtcacg 180
DB 244500aatctgtataaccattctcttaagcctccgtagtcctccaagctgcaaggagtcacg 180
|||||
QY 181tcactccttaacatctcggggagaaactcgtctattaataagagatgaaccccgtttt 240
|||||
```

Db 244560 tcactcctacatctcggggagaactcgcgtatttaataagaagatgaaccccgctcttt 244619

QY 241 agatccaagaacaagtgcgaactctttccagaatacagccttgccaacatactatgagg 300  
|||||

Db 244620 agatccaagaacaagtgcgaactctttccagaatacagccttgccaacatactatgagg 244679  
|||||

QY 301 attagttcgaagaataatctttccaggaaataagagcctgccttgcagaagaactactc 360  
|||||

Db 244680 attagttcgaagaataatctttccaggaaataagagcctgccttgcagaagaactactc 244739  
|||||

QY 361 tctttcctgcagcgaactcacttatactttcaactgaaatacagcttttggagtaattg 420  
|||||

Db 244740 tctttcctgcagcgaactcacttatactttcaactgaaatacagcttttggagtaattg 420  
|||||

QY 421 cgacccttaacagctgaagactttatagaatttggaaacaagttagctactccaagaagt 244799  
|||||

Db 244800 cgacccttaacagctgaagactttatagaatttggaaacaagttagctactccaagaagt 244859  
|||||

QY 481 ctcaaggaatctatgcttttgccttgaatccaatcaaaatgtacgaaagatccaagaggg 540  
|||||

Db 244860 ctcaaggaatctatgcttttgccttgaatccaatcaaaatgtacgaaagatccaagaggg 540  
|||||

QY 541 acaacctccatagacacatttgtagtgcaactcctcctaataatgtacactgttgtttac 600  
|||||

Db 244920 acaacctccatagacacatttgtagtgcaactcctcctaataatgtacactgttgtttac 244979  
|||||

QY 601 cctggaatcccaacctcgcgattcttcaaaccttttagctcttcagctctttccocgt 660  
|||||

Db 244980 cctggaatcccaacctcgcgattcttcaaaccttttagctcttcagctctttccocgt 245039  
|||||

QY 661 tcaataatctcaagaacacctgcgaattccaatactctacactatagaagcggagcttcta 720  
|||||

Db 245040 tcaataatctcaagaacacctgcgaattccaatactctacactatagaagcggagcttcta 245099  
|||||

QY 721 tctcaaaaataccaacaaacaaatggataaaactctcaaaaacctcactactataa 780  
|||||

Db 245100 tctcaaaaataccaacaaacaaatggataaaactctcaaaaacctcactactataa 245159  
|||||

QY 781 tcaaatgcaggtggaactaaacagattcacgttacccttcaatcccgatgcaaacacgc 840  
|||||

Db 245160 tcaaatgcaggtggaactaaacagattcacgttacccttcaatcccgatgcaaacacgc 840  
|||||

QY 841 agcaaaactatttaacccgggaataactcaattggcagaagcctcttggggagaacgat 900  
|||||

Db 245220 agcaaaactatttaacccgggaataactcaattggcagaagcctcttggggagaacgat 245279  
|||||

QY 901 tctcaagaacacctatccaatttacgtctaaagggcaactcacctctttgatgtgc 960  
|||||

Db 245280 tctcaagaacacctatccaatttacgtctaaagggcaactcacctctttgatgtgc 245339  
|||||

QY 961 aggaacctcagctcactccaattcaataataaattccccccaacaaatgaagcttag 1020  
|||||

Db 245340 aggaacctcagctcactccaattcaataataaattccccccaacaaatgaagcttag 245399  
|||||

QY 1021 agaagccttagcatcagccttagtaagaagaactctgtctcaactatattcttagccg 1080  
|||||

Db 245400 agaagccttagcatcagccttagtaagaagaactctgtctcaactatattcttagccg 245459  
|||||

QY 1081 tgcaaaatgpcgatcatctcctacactacaataattcatagctatcccgaaacataaaa 1140  
|||||

Db 245460 tgcaaaatgpcgatcatctcctacactacaataattcatagctatcccgaaacataaaa 245519  
|||||

QY 1141 acaagagatgcaacacgccaagcttaagctaaaataactctttaagaagccttagaaga 1200  
|||||

Db 245520 acaagagatgcaacacgccaagcttaagctaaaataactctttaagaagccttagaaga 1200  
|||||

QY 1201 actccaatctactgctaagaatctggaacatcttaattattcttcccgcttcccgctc 245579  
|||||

Db 245580 actccaatctactgctaagaatctggaacatcttaattattcttcccgcttcccgctc 245639  
|||||

QY 1261 agcaagttctttactagtcacacttatacagaaacagtggaagaagcttaggtctgc 1320  
|||||

Db 245640 agcaagttctttactagtcacacttatacagaaacagtggaagaagcttaggtctgc 245699  
|||||



polypeptides AA136754-137949. The polypeptides can be used as vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences can also be used to control growth of the microorganism. Chlamydia trachomatis is responsible for a large number of diseases, e.g. eye diseases such as conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion conjunctivitis; genital diseases such as nongonococcal urethritis; epididymitis, cervicitis, salpingitis, peritphatitis, Bartholinitis; pneumopathy in breast feeding infants; and venereal lymphogranulomatosis. The polypeptides of the invention may be of use in treating these diseases.

Query Match 25.0%; Score 449.6; DB 20; Length 1038602;  
Best Local Similarity 56.7%; Pred. No. 2.4e-115;  
Matches 976; Conservative 0; Mismatches 696; Indels 50; Gaps 6

1 acttccccctgctaactatgtcgcagataaatctgctcatgatattgcaggctcaggggag 60  
||||||| || | ||||| ||||| ||||| ||||| ||||| ||||| |||||  
252200 ACTTCCCTCTTCGCGATTATGTTCAGACACCGCTGCATGTATGCTGGGTTAGGAGAGA 252141

61 aaattttcaaaaaactctagtattcc--ggaaattcgtatatgcgaagatatacgt 117  
 - - - - - ||| - - - - - ||| - - - - - ||| - - - - - ||| - - - - - |||  
 252140 GACTGTTTTCGAATCGGACACACATGTTTTCAGCAGGTCATTCCATCGCGCAGATATCACT 252081

QY 118 ggaatctgtataccaaattctccttagctctcctgtagtctctcaaggctgcaaggaagtgc 177  
||||| | | | | | | | | | | | | | | |  
Db 252080 GGAATCTGCTTTCCTCCTAGCATTAGCAACTTCTGGATTTCAAAATCCTCCTTAAGC 252021

178 cagtcaactcctctacatctcggggagaactcgtctattaataataagatgaacccgttc 237  
QY  
25200 AACCCATCGGTCCTCCAGCTACTCACACAGTTGCTGTAAAGCTAAAGATGATCTCGCAC 251961  
Db

**QY**    238 tttagatccaagaacaagtgcgactcttttcagaaatcggccttgcacacatactatga 297  
       ||||| | | | | | | | | | | | | | | | | | |  
**Ddb** 251960 ATTGATCCTCGAGAGGTTCGCCTCTTCTCATATCAATTGGATTCATCATCTCTATGA 251901

QY 238 gggattagttcaagaaaataatcttcaggaaatatagagcctgcttgcagaagacta 357  
||||| || ||||| | - - - ||||| ||||| ||||| ||||| |||||  
DB 251900 AGGATTGGTACAAGAA ---ACTCCTTCGGAGAAGCTCTCCCGCTTTAGCGGAGAGTTT 251844

QY 358 cctctttcctcgacgactcaattatactttaaactgaatcagcttttggagtaa 417  
|| ||| || | ||||||||| || |||| ||||||||| ||  
Db 251843 CTCTTATCCGAGATATAAAAACATTATCTTTCAACTTGAAAAAGCCTTTGGAGCAA 251784

DQ 418 TGGGACCCCTTAACAGGTGAAGACTTTATAGAATCTCGGAACAAGTAGCACTCAAGA  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 251783 TGGAGATCCTTATACCGGCCTCATGATTTCCTTGCTCGGAATGATGTTACAAATCG 251724

470 agtccacgaggaaccatgccttcggcgaacacacaaatgacgaaagagacacaga 337  
 ||| | | | | | | | | | | | | | | |  
 Db 251723 TGTCGCTAGTATTATTTCTTTGGCCTTTTCCTATTGA----- 251683

Db 251684 -----CGTGAATAAGGATTCTGGATTTTTCCTCCAAAGATCATCACTCTGTAT 251694

Db 251633 CAATCTCCTCACACCACTCCACATTTCTAAAGCTGCTTACCGTCCCGGTATTTTATCC 251574

Db 251573 TGTGCATTGCGAGCATCAGATATGGAAGAAGAAAATCTTCCGATATCTACTGGAGC 251514

Db 251513 TTTTTCCTTAAAGAGAGAAAGACCGAAGATGCTTAAAGCTAGAGAGAGCCCTACTA 251454

Db 251453 CTATAATAAGACGAGGTAGCTGTACAGAGATCTGTATACATCATCTCTGATCAACA 251394

1321	QY	tatccctattgtcgaagaattgtctcttctccaagcaacctatttctcaggaactt	1380
245700	Db	tatccctattgtcgaagaattgtctctcctcaagcaacctatttctcaggaactt	245759
1381	QY	ctctttagctacagagatgggttcgcagactttgctgatctctatggcatttctcaagat	1440
245760	Db	ctctttagctacagagatgggttcgcagactttgctgatctctatggcatttctcaagat	245819
1441	QY	ctttgtttatccatcaggagttctctctttatgcatacaacataaggacttctctagaat	1500
245820	Db	ctttgtttatccatcaggagttctctctttatgcatacaacataaggacttctctagaat	245879
1501	QY	tctacaacaatagacaagaagacatcaacacaaaacgtctgggaattagttgtcgaagc	1560
245880	Db	tctacaacaatagacaagaagacatcaacacaaaacgtctgggaattagttgtcgaagc	245939
1561	QY	tctctttacctagagacctttcatattattgagcagatctacacagcagcatttcaatt	1620
245940	Db	tctctttacctagagacctttcatattattgagcagatctacacagcagcatttcaatt	245999
1621	QY	tgctatgaataaaaaatttctc-taattctaggagttctcaccacaggagttgtggacttcc	1679
246000	Db	tgctatgaataaaaaatttctc-taattctaggagttctcaccacaggagttgtggacttcc	246059
1680	QY	gttatctgaaggaataattagcacctcttttaattctgcatacttgcagaacttgcaagcgaact	1739
246060	Db	gttatctgaaggaataattagcacctcttttaattctgcatacttgcagaacttgcaagcgaact	246119
1740	QY	tatactaaactgggtgcctttgttgggcacctgtttctctctgactgctcttctctctcca	1799
246120	Db	tatactaaactgggtgcctttgttgggcacctgtttctctctgactgctcttctctctcca	246179
RESULT 3			
AAZ014425/C			
ID AAZ014425 standard; DNA; 1038602 BP.			

XX	Complete genome sequence of Chlamydia trachomatis.
DE	
XX	
XX	
KW	Vaccine: eye disease: conventional trachoma: nonendemic trachoma:
KW	paratrachoma: inclusion conjunctivitis; genital disease: peritriptitis:
KW	nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW	Bartholinitis; pneumopathy; venereal lymphogranulomatosis: ss.

OS Chlamydia trachomatis.  
XX  
PN W09928475-A2.

PD 10-JUN-1999.  
XX  
PF 27-NOV-1998; 98WO-IB01939.

PR 04-NOV-1998; 98US-0107077.  
PR 28-NOV-1997; 97FR-0015041.  
PR 17-DEC-1997; 97FR-0016034.

PA (GEST ) GENSET.  
XX  
PI Griffais R;

DR WPI; 1999-371125/31.  
XX  
PT Genome sequence of *Chlamydia trachomatis*

PS Claim 1; Page 373-656; 1755pp; English.  
XX  
CC The present sequence represents the complete genome of Chlamydia



10

Align seg 1/1 to: RAD20238 from: 1 to: 1799

```

1 MetArgLysIleSerValGlyIleCysIleThrIleLeuLeuSerLeuSe 17
101 ATGGCAAGATATACAGTGGGAATCTGTATCACCATTCTCTTAGCCTCTC 150
17 vAluValLeuGlnGlyCysIleGluSerSerHisSerSerThrSerArg 34
151 CGTAGTCTCCCAAGGCTGCAAGAGTCCAGTCACTCTCTACATCTCGGG 200
34 lyGluLeuAlaIleAsnIleArgAspGluProArgSerLeuAspProArg 50
201 GAGAACTCGCTATTATATATATATATATATATATATATATATATATAT 250
51 GlnValArgLeuLeuSerGlnIleSerLeuValIleHisIleIleGlu 67
251 CAAGTGGGACTTCTTTCAGAAATCAGCCCTTCTCAAAATATCTATGAGG 300
67 yIleuValGlnGluAsnLeuSerGlyAsnIleGluProAlaLeuAlaG 84
301 ATTAGTTCAGAAATATATATATATATATATATATATATATATATATAT 350
84 luAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLysLeuLys 100
351 AAGACTACTCTCTTCTCGGAGCGACTCACTTATATCTTTTAACTGAAA 400
101 SerAlaPheIlePheSerAsnGlySerProLeuThrAlaGluAspPheIle 117
401 TCAGCTTTTGGAGTAAATGGGACCCCTTAACAGCTGAAGACTTTATAGA 450
117 uSerTrpLysGlnValAlaThrGlnIleValSerGlyIleTyrAlaPheA 134
451 AICITGGAAACAGAGTACTCTCAGAAAGTCTCAGAAATCTAGCTTTTG 500
134 laLeuAsnProIleLysAsnValArgLysIleGlnGluGlyHisLeuSer 150
501 CCTTGAATCCAAATTAATAATGTACGAAAGATCCAAAGAGGACACCTCTCC 550
151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValIleTh 167
551 ATAGACCAATTTGGAGTGCACCTCTCTAATGAATCTACACTTTGTTAC 600
167 rIleuGlnSerProThrSerHisPheLeuLysLeuLeuAlaLeuProValP 184
601 CTGGAATCCCAACCTCCGACTCTCTTAAACCTTTTACCTCTCCAGTCT 650
184 hePheProValHisLysSerGlnArgThrLeuGlnSerLysSerLeuPro 200
651 TTTTCCCGTTCATAAATCTCAAGAACCTCGAATCCAAATCTCTACCT 700
201 IleAlaSerGlyAlaPheTyrProLysAsnIleLysGlnLysGlnTrpI 217
701 ATAGCAAGCGGAGCTTTCTATCTCAAAATATCAAAACAAACAAATGGAT 750
217 eLysLeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValGluThrL 234
751 AAAACTCTCAAAACAAACCTCCTACTATATCAAGTCAGGTGGAAACTA 800
234 yThrIleThrIleHisPheIleProAspAlaAsnThrAlaAlaLysLeu 250
801 AAACGATTACGATTCACTTCAATCCCGATGCAACACAGCAGCAAAACTA 850
251 PheAsnGlnGlyLysLeuAsnTrpGlnGlyProTrpGlyGluArgIle 267
851 TTTAATACAGGAAACCAATCAATGGCAAGACCTCTCTGGGAGAGACGAT 900
267 eProGlnGlnThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerP 284
901 TCTCAAGAAACCTATCCAAATTTACAGTCTAAGGGGCACTTACACTCTT 950
284 heAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhePro 300

```

```

951 TTGATGTCGCAGGAACCTCATGTGCTCACCTTCAATATCAATAAATCCCC 1000
301 LeuAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGlu 317
1001 CTCACAATATGAAGCTTAGAGAGCCCTTAGCATGACCTTAGATTAAGGA 1050
317 uAlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisL 334
1051 AGCTCTTGTCTCACTATATCTTAGGCCGTGCAAAACTGCCGATCATC 1100
334 euLeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMet 350
1101 TCCTACCTACAAATATTCATAGCTATCCGCAACATCAAAACACAGAGATG 1150
351 AlaGlnArgGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuGluG 367
1151 GCACACGCCAAGCTTAGCTTAAAGAACTCTTTAAAGAACTTTAGAAGA 1200
367 uLeuGlnIleThrAlaLysAspLeuGluHisLeuAsnLeuIlePhePro 384
1201 ACTCAATCACTGCTTAAGATCTCGAATCTTAATCTTATCTTTCCCG 1250
384 alSerSerSerAlaSerLeuLeuValGlnLeuIleArgGluGlnTrp 400
1251 TTTCTCTGTCAGCAAGTCTTTACTAGTCCAACTTATACGAGACAGTGG 1300
401 LysGluSerLeuGlyPheAlaIleProIleValGlyLysGluPheAlaLe 417
1301 AARGAAGTTTAGGTTCTGATCTATCTTATGTCGGAAGAAATTTGCTCT 1350
417 uLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGly 434
1351 TCTCAAGCAGACCTATCTTCAGGGAACCTCTCTTAGCTACAGAGGAT 1400
434 rpPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePheAlaTyr 450
1401 GGTTCGAGAGCTTTGCTGATCTCTATGCTATGCTAAGCATCTTTGCTTAT 1450
451 ProSerGlyValProProTyrAlaIleAsnHisLysAspPheLeuGluI 467
1451 CATCAGAGTCTCTCTTATGCAATCAACCATAGGACTTCTTAGAAT 1500
467 eLeuGlnAsnIleGluGlnGlnAspHisGlnLysArgSerGluLeuV 484
1501 TCTCAAAACATAGACAGACAGATACCAAAACGCTCGGAATAG 1550
484 alSerGlnAlaSerLeuTyrLeuGluThrPheHisIleIleGluProIle 500
1551 TGTGCAAGCTTCTCTTACCTAGAGACTTTCATATATTAGGCGGATC 1600
501 TyrHisAspAlaPheGlnPheAlaMetAsnLysLysLeuSerAsnLeuG 517
1601 TACCACAGCGCATTTCAATTTGCTATGAATAAAACTTTCTAATCTAGG 1650
517 yValSerProThrGlyValValAspPheArgTyrAlaLysGluAsn 532
1651 AGTCTCACCACAGAGGTTGTGACTTCCGTATGCTTAGCTAGGAAAT 1696

```

seq\_name: /SIDS5/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AA91990

seq\_documentation\_block:

ID AA91990 standard; DNA; 1230025 BP.

XX AA91990;

XX 13-SEP-1999 (first entry)

XX Nucleotide sequence of the complete genome of Chlamydia pneumoniae.

DE Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope; ss.

XX

OS Chlamydia pneumoniae.

XX PN W0927105-A2.

XX PD 03-JUN-1999.

XX PF 20-NOV-1998; 98WO-IB01890.

XX PR 04-NOV-1998; 98US-0107078.

XX PR 21-NOV-1997; 97FR-0014673.

XX PA (G8ST ) GENSET.

XX PI Griffais R;

XX PY WPI; 1999-357842/30.

XX PT Genome sequence of Chlamydia pneumoniae

XX PS Claim 1; Page 291-611; 1912pp; English.

XX The present sequence represents the complete genome of Chlamydia pneumoniae, and encodes proteins AY34584-Y35879. C. pneumoniae causes respiratory disease such as pneumonia and bronchitis and is thought to be a contributing factor in heart disease, sarcoidosis, sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The polypeptides encoded by the open reading frames of the C. pneumoniae genome (see AY34584-Y35879) can be used in immunogenic compositions as vaccines. Vectors containing C. pneumoniae nucleotide sequences can also be used as immunogenic compositions, especially where the vector directs the expression of a neutralising epitope of C. pneumoniae.

XX Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

## alignment\_scores:

Quality: 2726.00 Length: 533

Ratio: 5.124 Gaps: 1

Percent Similarity: 99.812 Percent Identity: 99.812

## alignment\_block:

US-09-824-567-2 x AAX91990 ..

Align seg 1/1 to: AAX91990 from: 1 to: 1230025

1 MetArgLysIleSerValGlyIleCysIleThrIleLeuLeuSerLeuSe 17

244480 ATGCGCAGATATCATGTCGGAACTCTGTATCACCATTCTCCTAGCCTCTC 244529

17 rValValLeuGlnGlyCysLysGluSerSerHisSerThrSerArgG 34

244530 CGTAGCTCTCCAGGCTGCAAGGAGTCCAGTCACTCCTCTATCATCTCGGG 244579

34 lylGluLeuAlaIleAsnIleArgAspGluProArgSerLeuAspProArg 50

244580 GAGAACTCGCTATTATATAGAGATGACCCCGCTCTTTAGATCCAGA 244629

51 GlnValArgLeuLeuSerGluIleSerLeuValLysHisIleTyrGluG 67

244630 CAAGTCGCACTCTCTTCAGAAATCAGCTGTCAACATATCTATGAGGG 244679

67 YLeuValGlnGluAsnLeuSerGlyAsnIleGluProAlaLeuAlaG 84

244680 ATTAGTTCAGAAATAATATCTTTTCAGAAATATAGAGCTCTCTTCAG 244729

84 luAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLysLeuLys 100

244730 AAGACTACTCTCTTCTCCGAGGACTCAGCTATACCTTTTAACTGAAA 244779

101 SerAlaPheTrpSerAsnGlyAspProLeuThrAlaGluAspPheIleG 117

244780 TCAGCTTTTGGAGTAATGCGACGCCCTTAACAGCTGAGACTTTATAGA 244829

117 uSerTrpLysGlnValAlaThrClnGluValSerGlyIleTyrAlaPheA 134  
244830 ATCTTGGAAACAGTAGTACTCAAGAACTCTCAGGAATCTATGCTTTTG 244879

134 laLeuAsnProIleLysAsnValArgLysIleGlnGluGlyHisLeuSer 150  
244880 CCTTGATCCCAATTAAATATGACAAAGATCCAGAGGAGACACCTCTCC 244929

151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValValTh 167  
244930 ATAGACCATTTTGGAGTGCACTCTCTTAATGAATCTACACTTGTGTTAC 244979

167 rLeuGluSerProThrSerHisPheLeuLysLeuAlaLeuProValP 184  
244980 CCTGGAATCCCCAACCTCGCATTTCTTAAACTTTTAGCTCTTCCAGTCT 245029

184 hePheProValHisLysSerGlnArgThrLeuGlnSerLysSerLeuPro 200  
245030 TTTTCCCGCTTCATAANTCTCAAGAACCTCGCAATCCAAATCTTACCT 245079

201 IleAlaSerGlyAlaPheTyrProLysAsnIleLysGlnLysGlnTrpIl 217  
245080 ATAGCAAGCGGAGCTTCTATCTCTTAAATAATCAACAAAAACATGGAT 245129

217 eLysLeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValGluThrL 234  
245130 AAAACTCTCAAAAAACCTCTACTATATAATCAAAAGTCAGGTGGAACATA 245179

234 ystThrIleThrIleHisPheIleProAspAlaAsnThrAlaAlaLysLeu 250  
245180 AAAGATTAGGATTCATCTTATCCCGATCGAACACACAGCAGCAAAACTA 245229

251 PheAsnGlnGlyLysLeuAsnTrpGlnGlyProTrpGlyGluArgIl 267  
245230 TTTAATCAGGAAACTCAATTTGCAAGGACCTCTTGGGAGAACGAT 245279

267 eProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerP 284  
245280 TCCTCAGAAACCTTATCCAAATTTACAGTCTAAGGGGCACTTACACTCTT 245329

284 heAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhePro 300  
245330 TTGATGTCCAGGAACTCATGGCTCACTTCAATATCAATAAATTCCTCC 245379

301 LeuAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGl 317  
245380 CTCACAAATATGAGCTTAGAAGCTTAGCATCAGCTTAGATAGGA 245429

317 uAlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisL 334  
245430 AGCTCTTGTCTCACTATATTTCTAGGCGCTGCATAAACTGCCGATCATC 245479

334 euLeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMet 350  
245480 TCCTACTACAAATATTCATAGCTATCCCGACACACAAAAACAAAGATG 245529

351 AlaGlnArgGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuGluGl 367  
245530 GCACACGCCAAGCTTACGCTAAAAAATCTTTAAAGAGCTTTAGAGA 245579

367 uLeuGlnIleThrAlaLysAspLeuGluHisLeuAsnLeuIlePheProV 384  
245580 ACTCCAAATACCCCTTAAGATCTCGAATCTTAAATCTTATCTTTCCCG 245629

384 alSerSerAlaSerSerLeuValGlnLeuIleArgGluGlnTrp 400  
245630 TTTTCTCTCAGCAGGTTCTTTACTAGTCCAACTTTATCGAGAACAGTGG 245679

401 LysGluSerLeuGlyPheAlaIleProIleValGlyLysGluPheAlaLe 417  
245680 AAGAAAGATTAGGTTCTGCTATCTCTATTCCTGTAAGGAATTCCTCT 245729

417 uLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGlyT 434

245730 TCTCCAGCAGACCTATCTTCAGGAACTCTCTTTAGCTACAGGAGAT 245779  
 434 tpPheAlaaspPheAlaaspProMetAlaPheLeuThrIlePheAlaTyr 450  
 245780 GGTTCGACAGCTTGTCTGATCTCTATGGCATTTCTAACGATCTTTGCTTAT 245829  
 451 ProSerGlyValProProTyrAlaIleAsnHisLysAspPheLeuGluI 467  
 245830 CCATCAGAGTCTCTCTATGCAATACCACTAAGGACTCTCTAGAAAT 245879  
 467 eLeuGlnAsnIleGluGlnGluAspHisGlnLysArgSerGluLeuV 484  
 245880 TCTACAAACATACAGACAGACAGATCACAAAAACGCTCGGAATAG 245929  
 484 alSerGlnAlaSerLeuTyrLeuGluThrPheHisIleIleGluProIle 500  
 245930 TGTCCCAAGCTCTCTCTTACCTAGAGACTTTCATATATTATGACGATC 245979  
 501 TyrHisAspAlaPheGlnPheAlaMetAsnLysLysLysLeuSer AsnLeuG 517  
 245980 TACACAGCGAGATTCATTTGCTATGATGATGATGATGATGATGATGAT 246029  
 517 lylValSerProThrGlyValValAspPheArgTyrAlaLysGluAsn 532  
 246030 GAGTCTCACCAACAGGAGTGTGGACTTCGGTTATGCTAAGGAAAT 246076

seq\_name: /SIDS5/gcdata/geneseq/geneseq-emb1/NA1999.DAT:AAZ01425

seq\_documentation\_block:

ID AAZ01425 standard; DNA; 1038602 BP.

XX AAZ01425;  
 AC AAZ01425;  
 XX 07-OCT-1999 (first entry)  
 XX Complete genome sequence of Chlamydia trachomatis.  
 XX Vaccine; eye disease; conjunctivitis; nonendemic trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihypertitis;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.  
 XX Chlamydia trachomatis.

XX W09928475-A2.

XX 10-JUN-1999.

XX 27-NOV-1998; 98WO-IB01939.

XX 04-NOV-1998; 98US-0107077.

XX 28-NOV-1997; 97FR-0015041.

XX 17-DEC-1997; 97FR-0016034.

XX (GEST ) GENSET.

XX Griffais R;

XX WPI; 1999-371125/31.

XX Genome sequence of Chlamydia trachomatis

XX Claim 1; Page 373-656; 1755pp; English.

XX The present sequence represents the complete genome of Chlamydia  
 CC trachomatis. Open reading frames (ORFs) of the genome encode  
 CC polypeptides AA36754-r37949. The polypeptides can be used as vaccines  
 CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also  
 CC be used to control growth of the microorganism. Chlamydia trachomatis is  
 CC responsible for a large number of diseases, e.g. eye diseases such as  
 CC conjunctival trachoma, nonendemic trachoma, paratrachoma, and inclusion  
 CC conjunctivitis; genital diseases such as nongonococcal urethritis,

CC epididymitis, cervicitis, salpingitis, perihypertitis, bartholinitis;  
 CC pneumopathy in breast feeding infants; and venereal  
 CC lymphogranulomatosis. The polypeptides of the invention may be of use in  
 CC treating these diseases.  
 XX  
 SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

alignment\_scores:

Quality: 1356.00 Length: 539

Ratio: 3.146 Gaps: 10

Percent Similarity: 79.963 Percent Identity: 50.278

alignment\_block:

US-09-824-567-2 x AAZ01425/rev ..

Align seg 1/1 to reverse of: AAZ01425 from: 1 to: 1038602

1 MetArgLysIleSerValGlyIleCysIleThrIleLeuLeuSerIleSe 17  
 252097 ATGCGCAAGATATCAGTGGGAATCTGCTGCTAGCATAGCAACTTC 252048  
 17 rValValLeuGlnGlyCysLysGluSerSerHisSerSerThrSerArg 34  
 252047 T.....GGATGTTCAAAATCTCTCTAAGCAACCAACATCGGT 252010  
 34 ly.....GluLeuAlaIleAsnIleArgAspGluProArgSer 46  
 252009 CTCACAGTACTCACACAGTGTGCTAAGCGTAAAGATGATCTCGACA 251960  
 47 LeuaspProArgGlnValArgLeuLeuSerGluIleSerLeuValIleHi 63  
 251959 TTTGATCTCGAGAGGTTCGCTTCTTTCTGATATCAATTTGATTCATCA 251910  
 63 sIleTyrGluGlyLeuValGlnGluAsnLeuSerGlyAsnIleGluP 80  
 251909 TCTCTATGAGGATGTGTACAAAGAACTCCT...TCTGGAGAACTCTCC 251863  
 80 rAlaLeuAlaGluAspTyrSerLeuSerSerAspGlyLeuThrIleThr 96  
 251862 CTGCTTACGCGAGAGTTCCTTATCCGAGATATAAAACTTATCT 251813  
 97 PheLysLeuLysSerAlaPheTrpSerAsnGlyAspProLeuThrAlaGl 113  
 251812 TTCACCTTGAAGAAAGCTTTTGGAGCAATGGAGATCTTATTACCGCTCA 251763  
 113 uaspPheIleGluSerTrpLysGlnValAlaThrGlnGluValSerGlyI 130  
 251762 TGATTTTCTGCTTCTGGAATGATGTGTACAAAAATCGTCTCGTAGTA 251713  
 130 leTyrAlaPheAlaLeuAsnProIleLysAsnValArgLysIleGlnGlu 146  
 251712 TTTATCTTTTCGCTTCTCCCTTCTCCCTATT...GACGTGAATTAAG..... 251675  
 147 GlyHisLeuSerIleAspHisPheGlyValHisSerProAsnGluSerTh 163  
 251674 .....GATTCTGGATTTTTCACAAAGATGATCATAC 251643  
 163 rLeuValValThrLeuGluSerProThrSerHisPheLeuLysLeuLeu 180  
 251642 TCTTGTATCAATCTCTCCTCACTCCAACTCCACATTTCTAAAGCTGCTTA 251593  
 180 laLeuProValPhePheProValHisLysSerGlnArgThrLeu..... 194  
 251592 CCCTCCCGTATTTTATCTCTGTGCAT...TCGACATCATCATATGAAA 251546  
 195 GlnSerLysSerLeuProIleAlaSerGlyAlaPheTyrProLysAsnI 211  
 251545 GAAGAAATCTCTCCGATCATCTCTGGAGCTTTTCTTAAAGAGAA 251496  
 211 eLysGlnLysGlnTrpIleLysLeuSerLysAsnProHisTyrTyrAsn 228  
 251495 GAAGACCGAGATGGTAAAGCTAGAGAGAGCCCTTACTACTATAATA 251446

```

annotation of identified clones (BLASTX, BLASTN and mapping to E.
coli K12 genome) please see firench-Constant et al. 2000, Nucleic
Acids Res.
Seq primer: M13 Forward
Class: shotgun.
FEATURES             Location/Qualifiers
     source            1..538
                        /organism="Photorhabdus luminescens"
                        /strain="W14"
                        /db_xref="taxon:29488"
                        /clone="PG01442"
                        /clone_lib="Photorhabdus luminescens strain W14 M13
                        library"
                        /dev_stage="primary phase variant"
                        /note="Genomic DNA from strain W14 was size selected (1-2
                        kb) and then cloned into M13 Janus."
BASE COUNT
150 a 111 c 129 g 148 t

```

```

alignment_scores:
    Quality: 148.50
    Ratio: 2.216
    Percent Similarity: 75.281
    Length: 89
    Gaps: 3
    Percent Identity: 38.202

alignment_block:
    US-09-824-567-2 x AQ990639 ..

Align seg 1/1 to: AQ990639 from: 1 to: 538

36  LeuAlaIleAsnAlleAArgSpGluProAArgSerLeuAspProAArgGlnVa 52
    |||:: |||:: ::|||::|||::|||::|||::|||::|||::|||::
246  TTTAGCTCGTAATAACGGCTCTGAACCGCAATCGTTAGATCCCATAGAT 295

52  largLeuLeuSerGluIleSerLeuValLysHisIleTyrGluGlyLeuV 69
    ::::: ::::: |||:: |||::|||::|||::|||::|||::|||::
296  TGAAGAGTGGCAAGATCAATTTGGCCCGTGATCTGTTGAAGGCTTGG 345

69  aISInGluAsnAsnLeuSerGlyVAsnIleGluProAlaLeuAlaGluAsp 85
    || ::::: |||::|||::|||::|||::|||::|||::|||::
346  TCATCGTTGGCCCG...ANTGGCGAAATTTGCCAGGTCCTGGACCAGT 392

86  TyrSerLeuSerSerAspGlyLeuThrTyrThrPhelLysLeu...LysSe 101
    ::::: ::::: |||::|||::|||::|||::|||::|||::|||::
393  TGGGAA...AATAAAGACTTTACCGTCTGGACATTCATATGGCTAAGA 439

101  rAlaPheTrpSerAsnGlyAspProLeuThrAlaGluAspPheIleGluS 118
    ::::: |||::|||::|||::|||::|||::|||::|||::|||::
440  TTCTAAGTGGCTAAATGGTGATCCGTGTCACGGCCCATTAITTTGTTTACA 489

118  eTrpLysGlnValAla 123
    |||::|||::|||::|||::|||::|||::|||::|||::
490  GTTGGCAACGCTGGCC 506

```

ORGANISM	Photobacterium luminescens. Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Photobacterium.
REFERENCE	1 (bases 1 to 538)
AUTHORS	french-Constant, R.H., Waterfield, N., Burland, V., Perna, N.T., Daborin, P.J., Bowen, D. and Blattner, F.R.
TITLE	A genomic sample sequence of the entomopathogenic bacterium Photobacterium luminescens W14: potential implications for virulence
JOURNAL	Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
MEDLINE	20378633
COMMENT	Contact: french-Constant RH Department of Biology and Biochemistry University of Bath South Building, Bath BA2 7AY, UK Tel: (44) 1225 826621 Fax: (44) 1225 826779 Email: bssrfc@bath.ac.uk This is one of 2,122 random reads from the M13 library. For

```

seq_name: qb_gss:BH370951

seq_documentation_block:
LOCUS
DEFINITION
  AG-ND-119H23.TF ND-TAM Anopheles gambiae genomic clone AG-ND-119H23
  , DNA sequence.
ACCESSION
  BH370951
VERSION
  BH370951.1 GI:17317076
KEYWORDS
  GSS.
SOURCE
  African malaria mosquito.
  Anopheles gambiae
  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
  Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea
  ; Anopheles.
  1 (bases 1 to 593)
AUTHORS
  Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE
  Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL
  Unpublished (2001)
COMMENT
  Other_GSSs: AG-ND-119H23.TR

```

Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjoftus@igrr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.

Seq primer: M13 For  
Seq primer: M13 Rev  
Class: BAC ends.

FEATURES  
source  
1..593  
/organism="Anopheles gambiae"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-119H23"  
/clone\_lib="ND-TAM"  
/note="Vector: pECBAC1; Site\_1: HindIII"  
BASE COUNT 97 a 180 c 171 g 145 t  
ORIGIN

alignment\_scores  
Quality: 137.00 Length: 159  
Ratio: 1.384 Gaps: 6  
Percent Similarity: 62.264 Percent Identity: 29.560

alignment\_block  
US-09-824-567-2 x BH370951/rev

Align seq 1/1 to reverse of: BH370951 from: 1 to: 593

291 TrpLeuThrPheAsnIle...AsnLysPheProLeuAsnAsnMetLysLeuA 307  
576 TATCTGCGCTTCAACACCAAGAAAGCGTTCGACACCGTGTGGTAC 527  
307 rgGluAlaLeuAlaSerAlaLeuAspLysGluAlaLeuValSerThrIle 323  
526 GCCAGGCTCTGAGCATCGCTGGACAAAGGCCATCGTTCGCGCGTA 477  
324 PheLeuGlyArgAlaLysThrAlaAspHisLeuLeuProThrAsnIleHi 340  
476 TTCAAGACACAGCGCACCGCGCTATTTCGATCTCGCGCGCGCATGCT 427  
340 sSerTyrProGluHisGlnLysGlnGluMetAlaGlnArgGlnAlaTyrA 357  
426 GGGCTATAACGAC...AAACTGCCCGAATACGCTTACACCCCGGAAAG 380  
357 lalysLysLeuPheLysGluAla...LeuGluGlnLeuGlnIleThrAla 372  
379 CGGCTGAACCTGCTCAAGCAGCGCGCGCTGGAGAAAGGGTTCGAGACC... 333  
373 LysAspLeuGluHisLeuAsnLeuIlePheProValSerSerAlaSe 389  
332 ...GATATCTGCTGATCGCGTAGCCGCCCTTACACCGCAAT...TC 289  
389 rSerLeuValGlnLeuIleArgGlnGluIntrLysGluSerLeuGlyP 406  
288 ACGCCCATCGCTCAGATCTCAAAACGACTGGCGGAA...GTGGCG 242  
406 heAlaIleProIleValGlyGluPheAlaLeuGlnAlaAspLeu 422  
241 TAAAGCCCAATCGTCACCTGGGAATGGGCCAGTATCTGGCGGGT 192  
423 SerSerGlyAsnPheSerLeuAlaThrGlyGlyTrpPheAlaAspPheAl 439

191 CGCAAGGGGGAACAGCAGCGCGCTGTATGCTGATGCTCGATAACGG 142  
439 aAspProMetAlaPheLeuThrIle 447  
141 CGACCGGACAACTTCGTACCCCTG 117

seq\_name: gb\_gss:B07758

seq\_documentation\_block:

LOCUS B07758 601 bp DNA linear GSS 15-NOV-1997  
DEFINITION 8219P101E01120195T7 Rhodobacter sphaeroides 2.4.1 genomic DNA  
library Rhodobacter sphaeroides genomic clone 8219P101E01120195T7  
similar to dppA (P23847), DNA sequence.

ACCESSION B07758  
VERSION B07758.1 GI:2055550  
KEYWORDS GSS,  
SOURCE Rhodobacter sphaeroides.  
ORGANISM Rhodobacter sphaeroides  
Bacteria: Proteobacteria; alpha subdivision; Rhodobacter group;  
Rhodobacter.

REFERENCE 1 (bases 1 to 601)

AUTHORS Choudhary,M., Mackenzie,C., Nereng,K., Sodergren,E., Weinstock,G.M.  
and Kaplan,S.

TITLE Low-resolution sequencing of Rhodobacter sphaeroides 2.4.1T:

Chromosome II is a true chromosome

JOURNAL Microbiology 143, 3085-3099 (1997)

MEDLINE 98015398

COMMENT Contact: Mackenzie, C.  
Department of Microbiology and Molecular Genetics  
The University of Texas Health Science Center  
6431 Fannin St., Houston, TX 77006, USA  
Tel: (713) 794-1742  
Fax: (713) 794-1782  
Email: mackenzie@ummg.med.uth.tmc.edu  
Seq primer: pBluescript SK (-) T7  
Class: shotgun.

FEATURES  
source

1..601  
/organism="Rhodobacter sphaeroides"  
/strain="2.4.1T"  
/db\_xref="taxon:1063"  
/clone="8219P101E01120195T7"  
/clone\_lib="Rhodobacter sphaeroides 2.4.1 genomic DNA  
library"  
/lab\_host="E. coli SL7-1"  
/note="Vector: pLA2917; Genomic DNA from Rhodobacter  
sphaeroides was prepared and partially digested with  
Sau3AI. Size selected (20kb) fragments were subcloned into  
the BglII site of the cosmid vector pLA2917 (Allen, L. N.  
and R. S. Hanson, 1985. Journal of Bacteriology,  
161:955-962. The library was then ordered around  
chromosome II of Rhodobacter sphaeroides (Choudhary et  
al., 1994. Journal of Bacteriology, 176:7694-7702). The  
cosmids were then digested with a variety of restriction  
enzymes (BamHI, EcoRI, BglII, PstI, EcoRV, NotI and  
DNaseI) and restriction fragments subcloned into the  
respective multiple cloning site sites of pBluescript SK  
(-). Note BglII fragments were subcloned into the  
pBluescript BamHI site. DNaseI fragments were subcloned  
into the EcoRV site. All subclones were transformed into  
E. coli XL1Blue MRF'. All fragments were then sequenced  
and the sequences where possible were assembled using the  
CGC program GELASSEMBLE."

BASE COUNT 101 a 203 c 175 g 108 t 14 others  
ORIGIN

alignment\_scores:

Quality: 124.50 Length: 160  
Ratio: 1.415 Gaps: 5  
Percent Similarity: 55.000 Percent Identity: 28.125



## alignment\_block:

US-09-824-567-2 x B07758/rev ..

Align seg 1/1 to reverse of: B07758 from: 1 to: 601

```

78 IleGluProAlaLeuAlaGluAspTyrSerLeuSerSerAspGlyLeuTh 94
534 ATCGCCCGGGTTCGGCGGAGATACGAGATCTCGGAGGACGGCTCAC 485
94 rTyrThrPheLysLeuLysSerAlaPhe...TrpSerAsnGlyAspProL 110
484 CTANACGTTCAAGTGGCCCGCGGCGTGAATTCACACGCGCGGAGA 435
110 euThrAlaGluAspPheIleGluSerTrpLysGlnValAlaThrGlnGlu 126
434 TGACGCGCGAGGATGTGAATATTCGTCGACCGCGGTGACCTGCTGG 385
127 ValSerGlyIleTyrAlaPheAlaLeuAsnProIleLysAsnValArgly 143
384 ACCCAGTCGCGCGCGGCGGCTTNTTCGCTCGATCAAGGGCTTCGATGC 335
143 sIleGlnGluGlyHisLeuSerIleAspHisPheGlyValHisSerProA 160
334 GATGCGCAGCGGTGCGCGCAGCAGCTCGAG...GGCGTGACGGTGGTGC 288
160 enGluSerThrLeuValThrLeuGluSerProThrSerHisPheLeu 176
287 ATCCCTCGACGTCAGGATCGAGCTCTGCGTCCCGACGCCACCTTCCTG 238
177 LysLeuLeuAlaLeuProValPhePheProValHisLysSerGlnArgTh 193
237 CATGTGATGGCGTGAANTTCGCTCGGTGGTGGCGGAAGGACGCTCGA 188
193 rLeuGlnSerLysSerLeu.....ProIleAlaSerGlyAlaPheT 207
187 GGCGCGCGCGCGGANTTCGCAACGCGGTCGCGCACCGCGGCTTC. 139
207 yrProLysAsnIleLysGlnLysGlnTrp.....IleLys 218
138 .....AAGCTCGCGGAATGACGCTGGCGCCAGCGCTCGTCTG 103
219 LeuSerLysAsnProHisTyrTyrAsnGln 228
102 TTCGAGAAGAACGCGACTACTGCGCGAG 73

```

seq\_name: gb\_gss:A0990866

seq\_documentation\_block:

```

LOCUS      A0990866                613 bp    DNA        linear    GSS 14-AUG-2000
DEFINITION Rf001703 Photorhabdus luminescens strain W14 M13 library
            Photorhabdus luminescens genomic clone PLC01703, DNA sequence.
ACCESSION  A0990866
VERSION    A0990866.1
KEYWORDS   GSS.
SOURCE     Photorhabdus luminescens.
           Photorhabdus luminescens.
           Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
           Photorhabdus.
REFERENCE  1 (bases 1 to 613)
AUTHORS   french-Constant,R.H., Waterfield,N., Burland,V., Perna,N.T.,
           Daborn,P.J., Bowen,D. and Blattner,F.R.
TITLE      A genomic sample sequence of the entomopathogenic bacterium
           Photorhabdus luminescens w14: potential implications for virulence
JOURNAL    Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
MEDLINE    20378633
COMMENT    Contact: french-Constant RH
           Department of Biology and Biochemistry
           University of Bath
           South Building, Bath BA2 7AY, UK
           Tel: (44) 1225 826621
           Fax: (44) 1225 826779
           Email: bssrf@bath.ac.uk

```

This is one of 2,122 random reads from the M13 library. For

annotation of identified clones (BLASTX, BLASTN and mapping to E. coli K12 genome) please see french-Constant et al. 2000, Nucleic Acids Res.

Seq primer: M13 Forward

Class: shotgun.

## FEATURES

```

source
  1..613
    /organism="Photorhabdus luminescens"
    /strain="W14"
    /db_xref="taxon:29488"
    /clone_lib="PLG01703"
    /clone="PLG01703"
    /dev_stage="primary phase variant"
    /notes="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus."

```

BASE COUNT 152 a 139 c 130 g 187 t 5 others  
ORIGIN

## alignment\_scores:

Quality: 122.50 Length: 241  
Ratio: 1.021 Gaps: 9  
Percent Similarity: 49.793 Percent Identity: 21.992

## alignment\_block:

US-09-824-567-2 x A0990866/rev ..

Align seg 1/1 to reverse of: A0990866 from: 1 to: 613

```

223 ProHisTyrTyrAsnGlnSerGlnValGluThrLysThrIleThi 239
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
606 CCAGNATATTACCGCTATCCAGTGAAGATAAATATACCAACAGCG 557
239 sPheIleProAspAlaAsnThrAlaAlaLysLeuPheAsnGlnGlyL 256
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
556 TTATTCCCGCAGCGCT..... 541
256 euAsnTrpGlnGlyProTrpGlyGluArgIleProGlnGluThrLeu 272
540 ..... 535
273 SerAsnLeuGlnSerLysGlyHisLeuHisSerPheAspValAlaGlyTh 289
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
534 AGAAATNAAGAAAGAGA.....TTCNTGATCAGGTTCCG 500
289 rSerTrpLeuThrPhe.....AsnIleAsnLysPheProLeuAsn.... 302
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
499 GTCAGTCTTACCTTTGTACGTTTATTACGAATTTACCAATAAAACCC 450
303 .....AsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAsp 315
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
449 GCGGTTTACTTGATGTCAGAGTCCGAGCGCTCAAGTTAAGTCFCGAT 400
316 LysGluAlaLeuValSerThrIlePheLeuGlyArgAlaLys..... 329
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
399 AGGGATTAATTATCCGGAATAATT...ATGGCAGCGGCAATACCGGCT 353
330 .....ThrAlaAspHisLeuLeuProThrAsnIleHisSerTyrProG 344
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
352 TATGGTTTACGCCAACATTTATTCGGGTGGCGCGGATTTTGTAAAGCCAG 303
344 luHisGlnLysGlnGluMetAlaGlnArgGlnAlaTyrAlaLysLysLeu 360
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
302 AATGGGCTAATTGGACAGGACAGCGCAATTAACGCCGAGAAATG 253
361 PheLysGluAla.....LeuGluLeuLeuGlnIleThrAlaLys 373
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
252 TTAGCTCAGCGCGGTTTCAATCAGCGCAATCCATTCAGATTCAACA.... 208
373 sAspLeuGluHisLeuAsnLeuIlePheProValSerSerAlaSerS 390
    :||| :||| :||| :||| :||| :||| :||| :||| :|||
207 .....TTGCTATACAAACCTCTGAGCAATAAATAC 177

```

```

390 erLeuLeuValGlnLeuLeuArgGluGlnTrpLysGluSerLeuGlyPhe 406
   ::      ::      ::      ::      ::      ::      ::      ::
176 AGCAGGCTATC...GCCGCCGATCCATCGTGGAAAGAGATTGGGAGC 130
   ::      ::      ::      ::      ::      ::      ::      ::
407 AlalleProfileValGlyLysGluPheAlaLeuLeuGlnAlaAspLeuSe 423
   ::      ::      ::      ::      ::      ::      ::      ::
129 GATGTGATTTTACAGATCAGCAATGGAACACCTCTTGAAGCGGCCA 80
   ::      ::      ::      ::      ::      ::      ::      ::
423 rSerGlyAsnPheSerLeuAlaThrGlyGlyTrpPheAlaAspPheAlaA 440
   ::      ::      ::      ::      ::      ::      ::      ::
79 TCAAGGGAATTATGAGTGACAGACCACTGGTGTGCTGATTACAATG 30
   ::      ::      ::      ::      ::      ::      ::      ::
440 sProMetAlaPheLeuThrile 447
   ::      ::      ::      ::      ::      ::      ::      ::
29 AACCTACAGCATTTTAAACATG 7

```

seq\_name: gb\_gss:AZ302752

seq\_documentation\_block:

LOCUS AZ302752 402 bp DNA linear GSS 06-MAR-2001  
DEFINITION GSSBU1690 Brucella abortus random genomic library Brucella  
melitensis biovar Abortus genomic clone U01690, DNA sequence.

ACCESSION AZ302752  
VERSION AZ302752.1 GI:10128963  
KEYWORDS GSS

SOURCE Brucella melitensis biovar Abortus.  
ORGANISM Brucella melitensis biovar Abortus

REFERENCE 1 (bases 1 to 402)  
AUTHORS Sanchez,D.O., Zandomeni,R.O., Cravero,S., Verdun,R.E., Pierrou,E.,  
Faccio,P., Diaz,G., Lanzavecchia,S., Aguero,F., Frasch,A.C.C.,  
Andersson,S.G.E., Rosetti,O.L., Grau,O. and Ugalde,R.A.

TITLE Gene discovery through genomic sequencing of Brucella abortus  
JOURNAL Infect. Immun. 69 (2), 865-868 (2001)  
COMMENT Contact: Siv Andersson  
Small Genomes Sequencing Group  
Department of Molecular Evolution, Uppsala University  
Norbyvagen 18C, S-752 36, Uppsala, Sweden  
Tel: 46-18-471-4379  
Fax: 46-18-471-6404  
Email: Siv.Andersson@ehc.uu.se

Sequences were basecalled with phred and vector was masked with  
crossmatch (see <http://genome.washington.edu>). Sequences were then  
trimmed from both ends to remove low quality bases and masked  
vector.  
Class: shotgun.

FEATURES Location/Qualifiers

source 1..402  
/organism="Brucella melitensis biovar Abortus"  
/strain="2308"  
/db\_xref="taxon:235"  
/clone="U01690"  
/clone\_lib="Brucella abortus random genomic library"  
/note="Vector: modified M13"

BASE COUNT 105 a 110 c 98 g 89 t

ORIGIN

alignment\_scores:

Quality: 118.00 Length: 130  
Ratio: 1.595 Gaps: 6  
Percent Similarity: 56.923 Percent Identity: 27.692

alignment\_block:

US-09-824-567-2 x AZ302752 ..

Align seq 1/1 to: AZ302752 from: 1 to: 402

```

220 SerLysAsnProHisTyrTyAsnGlnSerGlnValGluThrLysThrIL 236
   ::      ::      ::      ::      ::      ::      ::      ::
8 ACCCGCATGATAGCTAGTGGGGGCCAAGCGCTGCT...TATGAGCAGAT 54

```

```

236 eThrIleHisPheIleProAspAlaAsnThrAlaAlaLysLeuPheAsnG 253
   ::      ::      ::      ::      ::      ::      ::      ::
55 AACTCTCAGGTATACCCGATCCGATACACGGCGGATTCGCTCGAGG 104
   ::      ::      ::      ::      ::      ::      ::      ::

```

```

253 IngLysLeuAsn.....TrpGlnGly.....ProPro 262
   ::      ::      ::      ::      ::      ::      ::      ::
105 CTGGCGAAATAGACCTGATCTATGGAACGAGCGCGGATTCGCCCA 154
   ::      ::      ::      ::      ::      ::      ::      ::

```

```

263 TrpGlyGluArgIleProGlnGluThrLeuSerAsnLeuGlnSerLysgl 279
   ::      ::      ::      ::      ::      ::      ::      ::
155 TACTTTCGAGCGCTCCAGAA.....AATGG 180
   ::      ::      ::      ::      ::      ::      ::      ::

```

```

279 yHisLeu.....HisSerPhe AspValAlaGlyThrSerTrpLeuThr 293
   ::      ::      ::      ::      ::      ::      ::      ::
181 GCATCTATATACCGAGCTTCGGAGCCGTTAGAAACGTCGCTTTGGCG 230
   ::      ::      ::      ::      ::      ::      ::      ::

```

```

294 PheAsnIleAsnLysPheProLeuAsnAsnMetLysLeuArgGluAlaLe 310
   ::      ::      ::      ::      ::      ::      ::      ::
231 CTCAACACCAATCACGCGCAACAAAGACCTCGTCGCGCAAGCCAT 280
   ::      ::      ::      ::      ::      ::      ::      ::

```

```

310 uAlaSerAlaLeuAspLysGluAlaLeuValSerThrIlePheLeuGlyA 327
   ::      ::      ::      ::      ::      ::      ::      ::
281 CAATCATGCAGTCGATAGGACAGATGTTGCCACCGTCTCTACGGGA 330
   ::      ::      ::      ::      ::      ::      ::      ::

```

```

327 rgAlaLysThrAlaAspHisLeuLeuProThrAsnIle 339
   ::      ::      ::      ::      ::      ::      ::      ::
331 CACAGAAGCGCTGCAGATACCTGTTGCCGATATGTT 368
   ::      ::      ::      ::      ::      ::      ::      ::

```

seq\_name: gb\_gss:AQ012177

seq\_documentation\_block:

LOCUS AQ012177 713 bp DNA linear GSS 04-JUN-1998

DEFINITION 271IC073112697 Cosmid library of chromosome II Rhodobacter  
sphaeroides genomic clone 271IC073112697, DNA sequence.

ACCESSION AQ012177

VERSION AQ012177.1 GI:3177000

KEYWORDS GSS.

SOURCE Rhodobacter sphaeroides.

ORGANISM Rhodobacter sphaeroides

REFERENCE 1 (bases 1 to 713)

AUTHORS Choudhary,M., Mackenzie,C., Mouncey,N., Weinstock,G.M. and Kaplan

S.

TITLE RSGDB, the Rhodobacter sphaeroides Genome Database

JOURNAL Unpublished (1998)

COMMENT Contact: Choudhary, M.  
Department of Microbiology and Molecular Genetics  
University of Texas Medical School  
6431 Fannin Street, Houston, TX 77030, USA  
Tel: 713 500 5437  
Fax: 713 500 5499  
Email: madhu@utmsi.med.utmc.edu  
Seq primer: pbluescript T3  
Class: shotgun.

FEATURES Location/Qualifiers

source 1..713  
/organism="Rhodobacter sphaeroides"  
/strain="2.4.1r"  
/db\_xref="taxon:1063"  
/clone="271IC073112697"  
/clone\_lib="Cosmid library of chromosome II"  
/lab\_host="E. coli S17-1"  
/note="Vector: pLA2917"

BASE COUNT 132 a 250 c 216 g 115 t

ORIGIN

alignment\_scores:

Quality: 114.50 Length: 172  
Ratio: 1.180 Gaps: 5  
Percent Similarity: 56.395 Percent Identity: 23.256



67	GlyLeuValGlnGluAsnAsnLeuSerGlyAsnIleGluProAlaLeuA183
:	:
365	AAGCTGGTGCACGTGGATGCC...GAGGCGCAGATCATCCCATCTGGC411
:	:
83	aGUAspTyrSerLeuSerSeraspGlyLeuThrTyrThrPheLysLeu..99
:	:
412	CGAGCGCTGGACCGCTTCCTCGAGGATCAGAAAACCTACACGCTGTATCTGC461
:	:
100	..LysSerAlaPheTrpSerAsnGlyAspProLeuThrAlaGluAspPhe115
:	:
462	GCAAGGGCGCAAATTCACAGCAGCGCACGCCCTTCGATGCC.....502
:	:
116	IleGluSerTrpLysGln...ValAlaThrGlnGluValSerGlyIleTyr131
:	:
503	..AGCGCGTGGAATTCACCTGTGTGCGGGCAGAAGAACCTCCCTTTGGC549
:	:
131	rAlaPheAlaLeuAsnProIleLysAsnValArgLyIleGlnGluGlyH148
:	:
550	CCGCAATGAACCTGAATACGTACGACAGATCAGCGTGTGGATGACACA599
:	:
148	IsfSeuIleAsp152
:	:
600	CCAATTAAAGTTGAA613
:	:
seq_name: gb_gss:B07745	
seq_documentation_block:	
LOCUS	B07745 498 bp DNA linear
DEFINITION	8025R101A07082895T3 Rhodobacter sphaeroides 2.4.1 genome library Rhodobacter sphaeroides genomic clone 8025R101A07082895T3 similar to dppA (P23847), DNA sequence.
ACCESSION	B07745
VERSION	B07745.1 GI:2055537
KEYWORDS	GSS.
SOURCE	Rhodobacter sphaeroides.
ORGANISM	Rhodobacter sphaeroides. Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum rubrum.
REFERENCE	1 (bases 1 to 498) Choudhary,M., Mackenzie,C., Nereng,K., Sodergren,E., and Kaplan,S.
AUTHORS	Low-resolution sequencing of Rhodobacter sphaeroides chromosome II is a true chromosome
TITLE	Micobiology 143, 3085-3099 (1997)
JOURNAL	8015398
MEDLINE	Contact: Mackenzie, C.
COMMENT	Department of Microbiology and Molecular Genetics The University of Texas Health Science Center 6431 Fannin St., Houston, TX 77006, USA Tel: (713) 794-1742 Fax: (713) 794-1782 Email: mackenziet@utmsc.med.utah.tmc.edu Seq primer: pBluescript SK (-) T3 Class: shotgun.
FEATURES	Location/Qualifiers
source	1..498 /organism="Rhodobacter sphaeroides" /strain="2.4.1" /db_xref="taxon:1063" /clone="8025R101A07082895T3" /clone_lib="Rhodobacter sphaeroides 2.4.1 gss library" /lab_host="E. coli S17-1" /note="Vector: pLA2917; Genomic DNA from Rhodobacter sphaeroides was prepared and partially digested with Sau3AI. Size selected (20kb) fragments were ligated into the cosmid vector pLA2917 and R. S. Hanson, 1985. Journal of Bacteriology 161:955-962. The library was then ordered as chromosome II of Rhodobacter sphaeroides (Chromosome II). Journal of Bacteriology, 176:769-774, 1994. The library was then digested with a variety of restriction enzymes and sequenced."/> 

enzymes (BamHI, EcoRI, BglII, PstI, EcoRV, NotI and DnaSEI) and restriction fragments subcloned into the respective multiple cloning site sites of pBluescript SK (-). Note BglII fragments were subcloned into the pBluescript BamHI site. DnaSEI fragments were subcloned into the EcoRV site. All subclones were transformed into *E. coli* XLBlue MRF'. All fragments were then sequenced and the sequences where possible were assembled using the GCG program *GENASSEMBLE*.

BASE COUNT 83 a 137 c 171 g 88 t 19 others  
ORIGIN

alignment\_scores:  
Quality: 104.00 Length: 145  
Ratio: 1.268 Gaps: 7  
Percent Similarity: 56.552 Percent Identity: 30.345

alignment\_block:  
US-09-824-567-2 x B07745/rev ..

Align seg 1/1 to reverse of: B07745 from: 1 to: 498

```

93 LeuThrTyrThrPheLysLeuLysSerAlaPheTyrSerAsnGlyAspPr 109
|||||
362 CTCACCTACAGTTCCATCTGCGACGGAC.TTCTCCGCGCGCTCGCC 314
|||||
109 oLeuThrAlaGluAspPheLysLeuLysSerTyrPheValAlaThrGln 126
|||||
313 CATCACCCGAGGAC..... 298
|||||
126 luValSerGlyTyrAlaPheAlaLeuAsnProIleLysAsnValArg 142
|||||
297 .....GCGCCTTCCAGCTG.....CTG 280
|||||
143 LysIleGlnGluGlyHisLeuSerIle.....AspHisPheGlyValHi 157
|||||
279 CGCATCCGCGACGCAAGTCTCTGCTGCGGACACTATCGGTGAT 230
|||||
157 sSerProAsnGluSer.....ThrLeuValValThrLeuGluS 170
|||||
229 CGAACTNCCGAGCGCGACCGACCTCGTCTGCTCAAGCTGAAGA 180
|||||
170 eProThrSerHisPheLysLeuLysLeuAlaLeuProValPhePhePro 186
|||||
179 CCCCTCGCGCGCTTCTCTGCTGACCATGGCGATGCCGCGCTCGATC 130
|||||
187 ValHisLysSer.....GlnArgThrLeuGlnSerLy 197
|||||
129 CTGTCGAGGCGGCGCTCGAAGCCATGGCGGAGGAGGCTATCCGAGAA 80
|||||
197 sSerLeuProIleAlaSerGlyAlaPheTyr.ProLysAsnIleLysGln 213
|||||
79 G.....CCGCTGCGCTCGGCGCTTTACCGTCCAGGAATGCCGCC 36
|||||
214 LysGlnThrPheLysLeuSerLysAsnProHis 224
|||||
35 GCGCAGCCAGCTATCTCTAAANANAACCCGAAT 3
|||||
seq_name: gb_gss:TA380G06P

```

seq\_documentation\_block:  
LOCUS TA380G06P 519 bp DNA linear GSS 13-DEC-2000  
DEFINITION T. brucei sheared genomic DNA clone 380G06, forward sequence,  
genomic survey sequence.

ACCESSION AL497735 GI:11873457

VERSION AL497735

KEYWORDS GSS.

SOURCE Trypanosoma brucei.

ORGANISM Trypanosoma brucei

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

REFERENCE 1 (bases 1 to 519)

## AUTHORS

Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,  
Chillingworth, C., Omond, D., Harris, B., El-Sayed, N., Hou, L.,  
Meiville, S.E., Rajandream, M.A. and Barrell, B.G.

## TITLE

## JOURNAL

Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing  
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,  
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and  
nh@sanger.ac.uk

## COMMENT

Constructed at the Institute for Genomic Research (TIGR),  
Rockville, MD. Genomic DNA isolated from a cloned population of  
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared  
to give a tight size distribution (4 kb). The v + i method used for the library construction is  
described in detail in Smith, H. and Venter, J.C. (Making small  
insert libraries for whole genome shotgun sequencing projects. In  
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barrell, Oxford University Press, 1999).

Email: neilsayed@tigr.org  
Details of T. brucei sequencing at the Sanger Centre are available  
at [http://www.sanger.ac.uk/Projects/T\\_brucei/](http://www.sanger.ac.uk/Projects/T_brucei/).

## FEATURES

Location/Qualifiers

1..519

/organism="Trypanosoma brucei"

/strain="TREU927"

/db\_xref="taxon:5691"

/clone="380G06"

BASE COUNT 127 a 115 c 128 g 149 t

ORIGIN

alignment\_scores:

Quality: 103.50 Length: 143

Ratio: 1.344 Gaps: 4

Percent Similarity: 53.846 Percent Identity: 24.476

alignment\_block:

US-09-824-567-2 x TA380G06P ..

Align seg 1/1 to: TA380G06P from: 1 to: 519

250 LeuPheAsnGlnGly...LysLeuAsnTrpGlnGlyProTrpGlyG 265

|||||

110 TTAGAAATAGGGCGCACAAATGTTATGGAGCACCCGCCACAGTCGA 159

|||||

265 uArgIleProGln.....G 270

|||||

160 AAGGTTACCGCAGGTTTCGACGACGATGGAGGAGCTGCCAGCACCA 209

|||||

270 luThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerPheAsp 286

|||||

210 GTACACTCGAGCACTTCAGAGTTCTTCGTTCTTCACACATTTGATAC 259

|||||

287 AlaGlyThrSerTyrLeuThrPheAsnIleAsnLysPheProLeuAsn 303

|||||

260 TCAGGGCCATTCAGACCTTTCTCCGTC..... 289

|||||

303 nMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGluAlaLeu 320

|||||

290 .....CTCCTCTCATGTCGCAACATTTGGTGGATTCAGCGTAAGTC 332

|||||

320 alSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisLeuPro 336

|||||

333 AAACACATGTCATATGCGCAACAGCGGTGAATTTGAGCGTTTGGCG... 379

|||||

337 ThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMetAlaGln 353

|||||

380 .....CGTAGGTTGAAGACCTTCGCGGCACACTTGTCTCTCA 417

|||||

353 gGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuGluGln 370

|||||

418 GCAGAGTTACTTTTCAGGATCTGGAGAGTCGTAGATTACACACTCAGTG 467

|||||

370 leThrAlaLysAspLeuGluHisLeuAsn 379

```

468 CTTGACGAGGAGGCAAGGCAAGTAATCAAC 496
seq_name: gb_gss:BH398784
seq_documentation_block:
LOCUS      BH398784               554 bp    DNA          linear    GSS 11-DEC-2001
DEFINITION AG-ND-150B19, TR ND-TAM Anopheles gambiae genomic clone AG-ND-150B19
            , DNA sequence.
ACCESSION  BH398784
VERSION    BH398784.1  GI:17345000
KEYWORDS   GSS.
SOURCE     African malaria mosquito.
ORGANISM   Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea
            ; Anopheles.
REFERENCE  1  (bases 1 to 554)
AUTHORS    Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B.J.
TITLE      Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL    Unpublished (2001)
COMMENT    Other_GSSs: AG-ND-150B19, TF
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: b.loftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M University BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 Rev
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source            1..554
                        /organism="Anopheles gambiae"
                        /strain="PEST"
                        /db_xref="taxon:7165"
                        /clone="AG-ND-150B19"
                        /clone_lib="ND-TAM"
                        /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT  155 a 132 c 92 g 175 t
ORIGIN

alignment_scores:
  Quality: 103.50      Length: 176
  Ratio: 1.015         Gaps: 5
  Percent similarity: 57.955  Percent Identity: 20.455

alignment_block:
US-09-824-567-2 x BH398784/rev ..

Align seg 1/1 to reverse of: BH398784 from: 1 to: 554

290 SerTrpLeuThrPheAsn.....IleAsnLysPheProLe 301
   :::::  |||||
550 ACCTATACCATGTTCAATATGAGGATCCGATGTGGGGGGTATAGTCT 501
   ||:::
301 uAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGluA 318
   ||:::
500 AGACAAATAGCGCTACGCCGCGCATACCTTGGCATATACCAACAGG 451
   ||:::
318 laLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisLeu 334
   ||:::
450 AGCGGATTAAAGCACTTTATAAAGGTCAAGCGGTAGAGCTGAATCTTG 401

```

```

335 LeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMetal 351
   ::::|  :::::
400 GTGCCAGATGGGGTGCAGGGCAATGCTCAATATAGAGTAGTGTGTC 351
   :::::
351 aGlnArgGlnAlaTyrAlaLysLysLeu.....PheLysG 363
   |:::  ||:::
350 CTATAACCCGTTATTGGCCCAATAATTAAGTGGATCGATTGGTTATAAGA 301
   :::::
363 luAlaGluGluGluLeuGlnIleThrAlaLysAspLeuGluHisLeuAsn 379
   :::::
300 AAGGCAAGATGGTTATCGA...ACTTTACCAACAGCGGCAGCCATTACC 254
   |||
380 LeuIlePheProValSerSerAlaSerSerLeuLeuValGlnLeu1 396
   |||
253 CTGAATATATATCTGAAATAGCTCAAGCTCAGTCATTCATTCT..... 209
   |||
396 eArgGluGlnTrpLysGluSerLeu.....GlyPheAlaIlePro1 410
   |||
208 ...GAGTTATGGAAGAAAAAAGCTTAGATGCTATTGGGGTGGTGGTACT 163
   |||
410 leValGlyLysGluPheAlaLeuGlnAlaAspLeuSerSerGlyAsn 426
   :::::
162 TTAGGTCAGTAATTTTGGCGATAATTTAAAGAAAGCCATGCAATGTAA 113
   :::::
427 PheSerLeuAlaThrGlyTrpPheAlaAspPheAlaAspProMetal 443
   :::::
112 CATATGATTGGGGTGGCGCATGGATGCTGATTTTCTGAAGGAGAAAA 63
   :::::
443 aPheLeuThrIlePheAlaTyrProSer 452
   |||
62 TTTTGCTCAGTTGCTTTATGGTCCGAC 35

seq_name: gb_gss:TA358F03P
seq_documentation_block:
LOCUS      TA358F03P               558 bp    DNA          linear    GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 358F03, forward sequence,
            genomic survey sequence.
ACCESSION  AL494115
VERSION    AL494115.1  GI:11870744
KEYWORDS   GSS.
SOURCE     Trypanosoma brucei.
ORGANISM   Trypanosoma brucei
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma.
REFERENCE  1  (bases 1 to 558)
AUTHORS    Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
            Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
            Melville, S., Rajandream, M.A. and Barrell, B.G.
            Direct Submission
            Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
            project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
            Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
            nhles@sanger.ac.uk
            Constructed at the Institute for Genomic Research (TIGR),
            Rockville, MD. Genomic DNA isolated from a cloned population of
            Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
            to give a tight size distribution (
            4 kb). The v + i method used for the library construction is
            described in detail in Smith, H. and Venter, J.C. (Making small
            insert libraries for whole genome shotgun sequencing projects. In
            Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999).
            Email: nelsayed@tigr.org
            Details of T. brucei sequencing at the Sanger Centre are available
            at http://www.sanger.ac.uk/projects/T-brucei/.
FEATURES             Location/Qualifiers
     source            1..558
                        /organism="Trypanosoma brucei"
                        /strain="TREU927"
                        /db_xref="taxon:5691"
                        /clone="358f03"

```





**AUTHORS**  
**TITLE**  
**JOURNAL**  
**COMMENT**

Shetty J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B.J.  
 Direct Submission of BAC-end sequences from *Anopheles gambiae*  
 Unpublished (2001)  
 Other\_GSSS: AG-ND-174G16.TF  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: bjoftus@tigr.org

This clone is from an *A. gambiae* BAC library (ND-TAM) provided by  
 F.H. Collins and sequenced by The Institute for Genomic Research  
 (TIGR). The BAC library was generated from *A. gambiae* PEST strain  
 DNA. All DNA was extracted from newly hatched first instar larvae  
 to minimize the inclusion of DNA from microorganisms that inhabit  
 the gut. The DNA is derived from mixed sexes of larvae. The BAC  
 library was constructed at Texas A&M University BAC Center  
 University, College Station, Texas 77843-2123, USA using a HindIII  
 partial digest.  
 Seq primer: M13 Rev  
 Class: BAC ends.

**FEATURES**  
**Source**

Location/Qualifiers  
 1..583  
 /organism="Anopheles gambiae"  
 /strain="PEST"  
 /db\_xref="taxon:7165"  
 /clone="AG-ND-174G16"  
 /clone\_lib="ND-TAM"  
 /note="Vector: pECBAC1; Site\_1: HindIII"  
 BASE COUNT 161 a 133 c 99 g 190 t  
 ORIGIN

**alignment\_scores:**

Quality: 102.50 Length: 176  
 Ratio: 1.005 Gaps: 5  
 Percent Similarity: 57.955 Percent Identity: 20.455

**alignment\_block:**

US-09-824-567-2 x BH375296/rev ..

Align seq 1/1 to reverse of: BH375296 from: 1 to: 583

```

290 SerTrpLeuThrPheAsn.....IleAsnLysPheProLe 301
      ::::: ||||| ::::: ::::: ||
555 ACCTATACCATGTCATATCAAGGATCCGATTGTGGGGGGTATAGTCT 506
      ||:::|||||:::|||||:::|||||:::|||||:::|||||
301 uAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGluA 318
      ||:::|||||:::|||||:::|||||:::|||||:::|||||
505 AGACAAATAGCGGTACGCCGTCCTACCTTGGCATATATACCAACAGG 456
      ||:::|||||:::|||||:::|||||:::|||||:::|||||
318 laLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisLeu 334
      ::::: ::::: |||||:::|||||:::|||||:::|||||
455 AAGCGATTAAACACTTTATAAGGTCAAGCGTTAGAGCTGAATGTTG 406
      ||:::|||||:::|||||:::|||||:::|||||:::|||||
335 LeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMetAl 351
      ||:::|||||:::|||||:::|||||:::|||||:::|||||
405 GTGCCAGATGGGTGCAAGGGCACAATGCTCAATATAGAGTAGTGTTC 356
      ||:::|||||:::|||||:::|||||:::|||||:::|||||
351 aGlnArgGlnAlaTyrAlaLysLysLeu.....PheLysG 363
      ||::: |||||:::|||||:::|||||:::|||||
355 CTATACCCCGTTATTGGCCCAATAAATTACTGGATCGATTGGTTATAAGA 306
      ||::: |||||:::|||||:::|||||:::|||||
363 luAlaLeuGluLeuGlnIleThrAlaLysAspLeuGluHisLeuAsn 379
      ::::: ::::: |||||:::|||||:::|||||:::|||||
305 AGGCAAGAGTGTATCGA...ACTTTACCAACAGGGCAGCCATTAC 359
      ||:::|||||:::|||||:::|||||:::|||||:::|||||
380 LeuIlePheProValSerSerAlaSerLeuLeuValGlnLeuI 396
      ||| :::::|||||:::|||||:::|||||:::|||||
258 CTGAATATATATAGTAATAGCTCAAGCTCAGTCATTCATTCT.... 214
      ||:::|||||:::|||||:::|||||:::|||||
396 eArgGluGlnTrpLysGluSerLeu.....GlyPheAlaIleProI 410

```

```

213 ...GAGTTATCGAAGAAACCTAGATGCTATTGGGGTAAATCTTCAGT 168
      ||| |||||:::||||| ||| :::::
410 leValGlyLysGluPheAlaLeuLeuGlnAlaAspLeuSerSerGlyAsn 426
      ::: :::::|||||::: ||| :::::
167 TTAAGTCAGTAATTTTGGCGGATAATTTAAAGAAAGCCATGCAATGTAAA 118
      ||| |||||:::|||||::: |||||::: |||||
427 PheSerLeuAlaThrGlyGlyTrpPheAlaAspPheAlaAspProMetAl 443
      ::: |||||:::|||||::: |||||::: |||||
117 CATATGATTGGGGTGGGGCATGATGCTGATTTCTCTGAAGGAGAAAA 68
      ||| |||||:::|||||::: |||||::: |||||
443 aPheLeuThrIlePheAlaTyrProSer 452
      ||| ::::: |||||
67 TTTTCTCAGTTGCTTTATGTCGCGAAC 40

```

gb_ba:AF465631	+ 441.50	626.24	8771	: AF465631	Oenococcus oeni le
gb_ba:AE000265	- 441.00	611.58	36804	: AE000265	Enterococcus faec
gb_ba:AFI18935	- 440.50	601.51	3.6e-25	: AFI18935	Bacillus anthraci
gb_ba:ECU84242	+ 438.50	636.69	1.886	: U88424	Escherichia coli per
gb_ba:AE000231	+ 438.50	618.11	1.7990	: AE000231	Escherichia coli
seq_name:	gb_ba:AX349501				

```

-MODEL=framer_p2n.model -DEV=xlp
-O=/cgn2.1/USNP00.spool/USO9824567/runat_23072002_141350_13449
-DB=GenEmbl -QFMT=fastap -SUFFIX=rg -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOFCPI=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-NGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=bloms62 -TRANS=human40.cdi
-LIST=45 -DOCALL=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pts -NORM=ext -HEAPSIZ=500
-MINLEN=0 -MAXLEN=200000000 -USER=USO9824567 -CGN1=17895
-NCPU=3 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS=1

```

Query: US-09-824-567-2

Database: GenEmbl.\*

Database sequences: I/97656  
Database length: 1973333701

search time (sec): 2484.0900

score\_list:

gb\_pat:AX34

gb\_ba:AE001

gb\_ba:AP002

gb\_ba:AE002

qb\_ba:AE002:

gb\_pat:AX301

gb\_ba:AE001:

gb\_pat:AX02

gb\_ba:BSOPPO

qab\_ba:BGA250

ab ba:AF0430

g-b ha: AFOO-

gh\_ba:gsncr

95-046: 8A341

96\_yb\_da:KME60

gb\_ba:AE009

gb\_ba:AL5961

gb\_ba:AF0009

gb\_ba:AF1037

gb\_ba:S75873

gb\_ba:AF3053

```

|||||
401 CCTTGAATCCAAATTAATAATGTCAGAAAGATCCAGAGGGACACCTCTCC 450
151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValValTh 167
|||||
451 ATAGACCAATTTTGAGTGCACCTCTCTAATGAATCTACACTTGTGTTC 500
167 rLeuGluSerProThrSerHisPheLeuLysLeuAlaLeuProValp 184
|||||
501 COTGGAATCCCAACCTCGCAATTTCTTAAACTTTTAGCTCTTCCAGTCT 550
184 hePheProValHisLysSerGlnArgThrLeuGlnSerLysSerLeuPro 200
|||||
551 TTTTCCCGTTCATAAATCTCAAGAACCTGCAATCCAAATCTTACCT 600
201 IleAlaSerGlyAlaPheTyrProLysAsnLysGlnLysGlnPrpI 217
|||||
601 ATAGCAAGCGAGCTTCTATCTCTTAAATAATCAACAACAAATGGAT 650
217 eLysLeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValGluThrL 234
|||||
651 AAAACTCTCAAAAACCCCTCACTACTATAATCAAAAGTCAGGTGCAAACTA 700
234 ysThrIleThrIleHisPheIleProAspAlaAsnThrAlaAlaLysLeu 250
|||||
701 AAACGATTAGATTCTCTTATCTCCGATGCAAAACACAGCAGCAAACTA 750
251 PheAsnGlnGlyLysLeuAsnTrpGlnGlyProProTrpGlyGluArgI 267
|||||
751 TTATATCAGGAAACTCAATTTGCAAGGACCTCTTTGGGAGAACGAT 800
267 eProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerP 284
|||||
801 TCCTCAAGAACCTTATCAATTTAGCTTAAGGGCACTTACACTCTT 850
284 heAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhePro 300
|||||
851 TTGATGTCGAGGAACCTCATGGCTCACTTCAATATCAATAAATTCGCC 900
301 LeuAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysG 317
|||||
901 CTCACAATATGAAGCTTAGAGAGCCTTAGCATCAGCCTTAGAAGA 950
317 uAlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisL 334
|||||
951 AGCTTTGTCTCAACTATATTTCTTAGCCGCTGCAAAACTGCCGATCATC 1000
334 euLeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMet 350
|||||
1001 TCCTACTACAATATTCTAGCTATCCCAACATCAAAACCAAGAGATG 1050
351 AlaGlnArgGluAlaTyrAlaLysLysLeuPheLysGluAlaLeuGluCl 367
|||||
1051 GCACAACGGCAAGCTTAGCTTAAAAACTCTTTAAAGAGCTTTAGAAGA 1100
367 uLeuGlnIleThrAlaLysAspLeuGluHisLeuAsnLeuIlePhePro 384
|||||
1101 ACTCAAAATCACTGTAAGATCTCGAATCTTAATCTTACTTCTTCGCG 1150
384 alSerSerAlaSerSerLeuLeuValGlnLeuIleArgGluGlnTrp 400
|||||
1151 TTTCTCTGTCACCAAGTTCTTTACTAGTCCAACTTATAGCAACAGTGG 1200
401 LysGluSerLeuGlyPheAlaIleProIleValGlyLysGluPheAlaLe 417
|||||
1201 AAAGAAAGTTTAGGCTTCGCTATCCCTATTGTCGGAAGGAATTTGCTCT 1250
417 uLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGly 434
|||||
1251 TCTCCAGCAGACCTATCTTCAAGGAACCTCTCTTAGCTACAGGAGAT 1300
434 rpPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePheAlaTyr
|||||

```

```

1301 GGTTCGAGACTTTGCTGATCTATGCAATTTCTAACGATCTTTCTCTTAT 1350
451 ProSerGlyValProTyrAlaIleAsnHisLysAspPheLeuGluI 467
|||||
1351 CCATCAGGAGTTCTCTCTTATGCAATCAACCATAGGACTTCTCTAGAAAT 1400
467 eLeuGlnAsnIleGluGlnGluAspHisGlnLysArgSerGluLeuV 484
|||||
1401 TCTACAAAACATAGAACAAAGACAGATCACCAAAAACGCTCGGAATTAG 1450
484 alSerGlnAlaSerLeuTyrLeuGluThrPheHisIleIleLeuProIle 500
|||||
1451 TGTGCGAAGCTTCTCTTACCTAGGACCTTTCATATTATGAGCCGATC 1500
501 TyrHisAspAlaPheGlnPheAlaMetAsnLysLysLeuSerAsnLeuG 517
|||||
1501 TACCAGCGCATTTCAATTTGCTATGATGAATAAAAACCTTCTTAATCTAG 1550
517 yValSerProThrGlyValValAspPheArgTyrAlaLysGluAsn 532
|||||
1551 AGTCTACCCACAGAGGTGTGGACTTCGCTTATGCTAAGGAAAT 1596
seq_name: gb_pat:AX268341
seq_documentation_block:
LOCUS AX268341 1799 bp DNA linear PAT 29-OCT-2001
DEFINITION Sequence 1 from Patent WO0174863.
ACCESSION AX268341
VERSION AX268341.1 GI:16541562
KEYWORDS
SOURCE
ORGANISM Chlamydomophila pneumoniae.
REFERENCE
AUTHORS Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
TITLE Chlamydia antigens and corresponding dna fragments and uses thereof
JOURNAL Patent: WO 0174863-A 1 11-OCT-2001,
Aventis Pasteur Limited (CA)
FEATURES
source
location/Qualifiers
1..1799
/organism="Chlamydomophila pneumoniae"
/db_xref="taxon:83558"
101..1699
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAD10302.1"
/db_xref="GI:16541563"
/translation="MBKISVGICITILLISLVVLOGCKESSHSTSRGELAINIRDEP
RSLDPROVRLLSISLVKHYEGLVQENNLSGNIEPALAEDYSLSDGLTYTFKLKSA
FWSNGDPLTAEDEFESKQVATQEVSGIYAFALNPKNRKIQEHLSDHFGVHSPN
ESTLVLTSEPTSHFLKLLALEVFFVHKQSRTLOKSLPIASGAFYPKNIKQKOWIK
LSKNPHYNQSVETKTIITHTIPDNTAAKLNQKLNQWQPPMGERIPQETLSNLO
SKGHLHSFDVAGTSWLTFFNKFPLNNKRLREALSALDKALYSTIFELGAKTADHL
LPTNIHSYPEHQKQENQAQYAKKLFKEALELQITAKLEHLNLIFFVSSASSLL
VQLIREQKESLGFALPIVKGFEALLQADLSGNFSLATGGMFADFADPMAPLTFAY
PSGVPPIAINHDFLEILQNIQEQDQKRSLSQASLYLETFHTIPIYHDAFQFA
MNKLSNLGVSPTVDFRYAKEN"
BASE COUNT 560 a 439 c 294 g 506 t
ORIGIN
alignment_scores:
Quality: 2739.00 Length: 532
Ratio: 5.148 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-824-567-2 x AX268341 ..
Align seg 1/1 to: AX268341 from: 1 to: 1799
1 MetArgLysIleSerValGlyIleCysIleThrIleLeuLeuSerLeuSe 17

```

```

101 ATGCCCAAGATATCAGTGGAAATCTGTATCACCATCTCTTACGCTCTC 150
17 rValValLeuGlnGlyCysLysGluSerSerHisSerThrSerArg 34
151 CGTAGCTCTCCAGGCTGCAAGGATCCAGTCACTCTCTACATCTCGGG 200
34 yGluLeuAlaIleAsnIleArgAspGluProArgSerLeuAspProArg 50
201 GAGAACTCGCTATTATATATAAGAGATGAACCCCGTCTTTAGATCCAAGA 250
51 GlnValArgLeuSerGluIleSerLeuValLysHisIleThrGlu 67
251 CAAGTGCACCTCTCTTCAGAAATCAGCCTTGCAACATATCTATCAGGG 300
67 yLeuValGlnGluAsnAsnLeuSerGlyAsnIleGluProAlaLeuAlaG 84
301 ATTAGTTCAAGAAATAATCTTTTCAGGAATATAGAGCTGCTCTTCGAG 350
84 luAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLysLys 100
351 AAGACTACTCTCTCTCTCGGACGGACTCACTTATCTTTTAACTGAAA 400
101 SerAlaPheTrpSerAsnGlyAspProLeuThrAlaGluAspPheIle 117
401 TCAGCTTTTGGAGTAATGGCGGCCCTTACAGCTCAAGACTTTATAGA 450
117 uSerTrpLysGlnValAlaThrGlnGluValSerGlyIleTyrAlaPheA 134
451 ATCTGTGAAACAAAGTAGTACTCAAGAAGTCTCAGGAATCTATGCTTTG 500
134 laLeuAsnProIleLysAsnValArgLysIleGlnGluGlyHisLeuSer 150
501 CCTTGATTCATTAATAATGATGACAAAGATCCAAAGAGGACACCTCTCC 550
151 lleAspHisPheGlyValHisSerProAsnGluSerThrLeuValTh 167
551 ATAGACCATTTTGGAGTGCACCTCTCTTAATGAATCTACACTTGTGTAC 600
167 rLeuGluSerProThrSerHisPheLeuLysLeuAlaLeuProValP 184
601 CCTGGATCCCCAACCTCGCATTTCTTTAAACTTTTAGCTCTTCCAGTCT 650
184 hePheProValHisLysSerGlnArgThrLeuGlnSerLysSerLeuPro 200
651 TTTTCCCGCTGTCATAAATCTCAAGACCCCTGCAATCCAAATCTCTACCT 700
201 lleAlaSerGlyIlePheTyrProLysAsnIleLysGlnLysGlnTrpI 217
701 ATAGCAGGGAGGCTTTCTATCTCTAAAAATATCAAAACAAACATGGAT 750
217 eLysLeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValGluThrL 234
751 AAACCTCTCAAAAACCTCTACTACTATAATCAAACTCAGGTGGAACATA 800
234 ysthrIleThrIleHisPheIleProAspAlaAsnThrAlaAlaLysLeu 250
801 AAACGATTACGATTCACTTCTTCCCATGCAACACAGCAGCAAAACTA 850
251 pheAsnGlnGlyLysLeuAsnTrpGlnGlyProProTrpGlyGluArgI 267
851 TTTAATCAGGAAAACTCAATTTGCAAGGACCTCTCTTGGGAGAACCAT 900
267 eProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerP 284
901 TCTCTAAGAAACCTTATCCAAATTTACAGTCTAAGGGCACCTTACACTCT 950
284 heAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhePro 300
951 TTGATGTCCAGAACCTCATGCTCCTCACTTCAATATCAATAAATTTCCC 1000
301 LeuAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysG 317

```

```

1001 CTCACAATATGAAGCTTAGAGAAGCCCTTAGCATCAGCCTTAGATAAGA 1050
317 uAlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisL 334
1051 AGCTCTGTCTCAACTATATCTTAGGCCGTGCAAAACTGCCGATCATC 1100
334 euLeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMet 350
1101 TCTACTACTACAATATTCATAGCTATCCCGAATCAAAACAAAGAGATG 1150
351 AlaGlnArgGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuGluG 367
1151 GCACAACGCCAAGCTTAGCTTAAAAAACCTTTTAAAGAAGCTTTAGAGA 1200
367 uLeuGlnIleThrAlaLysAspLeuGluHisIleAsnLeuIlePhePro 384
1201 ACTCCAAATCACTGCTAAAGATCTCGAATCTTAATCTTATCTTCTCCG 1250
384 alSerSerAlaSerSerLeuLeuValGlnLeuIleArgGluGlnTrp 400
1251 TTTCTCTCGTCAGCAAGTCTTTACTAGTCCAACTTATACGAGAACAGTGG 1300
401 LysGluSerLeuGlyPheAlaIleProIleValGlyLysGluPheAlaLe 417
1301 AAAGAAGTTTAGGGTTGCTATCCCTATTGTCGGAAGGAATTTGCTCT 1350
417 uLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGly 434
1351 TCTCCAGCAGACCTTCTTTCAGGGAACCTCTTTTAGCTACAGGAGAT 1400
434 rpPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePheAlaTyr 450
1401 GGTTCGACAGCTTTGCTGATCCTTATGGCATTTCTAACGATCTTTGCTTAT 1450
451 ProSerGlyValProProTyrAlaIleAsnHisLysAspPheLeuGluI 467
1451 CCATCAGAGTTCTCTCTTATGCAATCAACCAAGGACTTCTCTAGAAAT 1500
467 eLeuGlnAsnIleGluGlnGlnAspHisGlnLysArgSerGluLeuV 484
1501 TCTCAAAACATAGAACACAGCAGATCCCAAAACGCTCGGAATTAG 1550
484 alSerGlnAlaSerLeuTyrLeuGluThrPheHisIleIleGluProIle 500
1551 TGTCCAAAGCTTCTCTTACCTAGAGACCTTCTATATTATTAGCCGATC 1600
501 TyrHisAspAlaPheGlnPheAlaMetAsnLysLysLeuSerAsnLeuG 517
1601 TACCACGACGATTTCAATTTGCTATGATATAAAACCTTTCTATCTAGG 1650
517 yValSerProThrGlyValValAspPheArgTyrAlaLysGluAsn 532
1651 AGTCTCACCAACAGGAGTTGTGACTTCGTTATGCTTAGGAAAAAT 1696

```

seq\_name: gb\_ba:AE001606

seq\_documentation\_block:

LOCUS AE001606

DEFINITION Chlamydia pneumoniae section 22 of 103 of the complete genome.

ACCESSION AE001606 AE001363

VERSION AE001606.1 GI:4376464

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE





```

/db_xref="GI:4376472"
/translation="MDNYLLNLTITSTNPKRTLIENLSLQKENRNIALVGSGL
GKTTIRAILGFUPENCLIKTGSILFDITIKLSPXELHKIRGQKIA"ILQNAWGL
TPSMRIQMIETIRGHRHMKKEEYKAMQLITDVCIPNPKYSFSQYFELSGGMRO
RVVIAIALASPKLILADDEPTTALDSMAOVLRLNIHQOQKOATILLVTHNLSLVK
ELCNDICIIKDGKLIETGTVEIFILSPKHPYTLKLLNANVSKIPFKTSSPILKNKFPQ
LMSWQGL"
gene 10578..11366
/feature="CDS"
/feature="gene"
/feature="note"
/feature="gene"
/feature="codon_start"
/feature="transl_table"
/feature="product"
/feature="protein_id"
/db_xref="GI:4376473"
/translation="MTTLISKDLIRGKILNHLNKLKGSVLIVGPSGSKS
SLAATILDLKPTTGTITPHMDPKIPRARKVQVITODIDSSLNPCMSIKGIISPLNI
IGTYSKAEONKEYNVLDVNLPKSVLHLPYKLSGQKRIARAKALVSPELLICD
EPSSLDTLNQSLIDLFQTKKETQNTLFTIHDMSAAYIADTIAVMDGSLVEHA
CREKIFSPKHTTQDLDAIPFISLISSTEMEPSEYELQVASK"
BASE COUNT 3743 a 2799 c 1831 g 3275 t
ORIGIN

```

```

alignment_scores:
  Quality: 2739.00      Length: 532
  Ratio: 5.148          Gaps: 0
  Percent Similarity: 100.000      Percent Identity: 100.000

alignment_block:
  US-09-824-567-2 x AE001606 ..

Align seg 1/1 to: AE001606 from: 1 to: 11648

1 MetArgLysIleSerValGlyIleCysIleThrIleLeuLeuSerLeuSe 17
|||||
1053 ATGGCGCAAGATATCAGTGGGAATCTGTATCACCATTCTCCTTAGCCTCTC 1102

17 rValValIleGlnGlyCysLysLysGluSerSerHisSerThrSerArgg 34
|||||
1103 CGTAGTCCTCCAAAGCTGCAAGGAGTCAGTCAGTCTCTCTATCATCTCGG 1152

34 lyGluLeuAlaIleAsnIleArgAspGluProArgSerLeuAspProArg 50
|||||
1153 GAGAACTCGCTATTAATATAGAGATCAACCCGCTCTTTAGATCCAAAGA 1202

51 GlnValArgLeuSerGluIleSerLeuValLysHisIleTyrGluG1 67
|||||
1203 CAAGTCGCACTCTCTTTCAGAAATCAGCCTTGTCAACACATATCTATGAGG 1252

67 yLeuValGlnGluAsnAsnLeuSerGlyAsnIleGluProAlaLeuAla 84
|||||
1253 ATTAGTTCAGAAATATATCTTTTCAGAAATATAGAGCTGCTCTTCGAG 1302

84 luAspTyrSerLeuSerSerAspGlyLeuThrThrPheLysLeuLys 100
|||||
1303 AAGACTACTCTCTTTCTCGGAGGACTACTATTATCTTTTAACTGAAA 1352

101 SerAlaPheTrpSerAsnGlyAspProLeuThrAlaGluAspPheIleG1 117
|||||
1353 TCAGCTTTTGGAGTAAATGCGACCCCTTAAACAGCTGAAGACTTTATAGA 1402

117 uSerTrpLysGlnValAlaThrGlnGluValLysGlyIleTyrAlaPheA 134
|||||
1403 ATCTTGAACCAAGTAGCTACTCAAGAAGTCTCAGGAATCTATGCTTTTG 1452

134 laLeuAsnProIleLysAsnValArgLysIleGlnGluGlyHisLeuSer 150
|||||
1453 CCITGATCCCAATTAATAATGTCAGAAAGATCCAAAGAGGAGACCTCTCC 1502

151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValTh 167
|||||

```

```

|||||
1503 ATAGACCAATTTGGAGTGCATCTCTCTTAATGAATCTACACTTGTGTAC 1552

167 rLeuGluSerProThrSerHisPheLeuLysLeuLeuAlaLeuProValP 184
|||||
1553 CTTGGAATCCCCCACTCGGATTTCTTAAAACTTTTAGCTCTTCCAGTCT 1602

184 hePheProValHisLysSerGlnArgThrLeuGlnSerLysSerLeuPro 200
|||||
1603 TTTTCCCGTTCATAAATCTCAAGAACCCTGCAATTCGAAATCTCTACCT 1652

201 lleAlaSerGlyAlaPheTyrProLysAsnIleLysGlnLysGlnTrpI1 217
|||||
1653 ATAGCAAGCGGAGCTTTCTATCTCTAAATAATCAAAACAAACAAATGGAT 1702

217 eLysLeuSerLysAsnProHisTyrTrpAsnGlnSerGlnValGluThrL 234
|||||
1703 AAACTCTCAAAAACCTCCTCAGTACTATTAATCAAAAGTCAGGTGAAACTA 1752

234 yThrIleThrIleHisPheIleProAspAlaAsnThrAlaAlaLysLeu 250
|||||
1753 AAACGATTACGATTCACTTCATTCGGATGCAACACACAGCAGCAAACTA 1802

251 PheAsnGlnGlyLysLeuAsnTrpGlnGlyProTrpGlyGluArgI1 267
|||||
1803 TTTAATCAGGAAAACTCAATTGGCAAGGACCTCCTTGGGAGAACGCAT 1852

267 eProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerP 284
|||||
1853 TCCTCAAGAAACCTTATCCATTTACAGTCATGAGGCGACTTACACTCTT 1902

284 heAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhePro 300
|||||
1903 TTGATGTCGAGGAACCTCATGCTCAGCTTCAATATCAATAAATTCCTCC 1952

301 LeuAsnMetLysLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysG1 317
|||||
1953 CTCACAAATATGAGCTTTAGAGAAGCCTTAGCATCAGCCTTAGATAGAAG 2002

317 uAlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisL 334
|||||
2003 AGCTCTGCTCAACATATATCTTAGCGCGTGCAGAAACCTGCCGATCATC 2052

334 euLeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMet 350
|||||
2053 TCCATCCTTACAAATATTCATAGCTATCCGAAACATCAAAACAAAGAGATG 2102

351 AlaGlnArgGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuGluG1 367
|||||
2103 GCACACGCCAAGCTTACGCTAAAAAACTCTTTAAAGAGCTTTAGAGA 2152

367 uLeuGlnIleThrAlaLysAspLeuGluHisLeuAsnLeuIlePheProV 384
|||||
2153 ACTCCAAATCACTGCTTAAGATCTCGACATCTTAATCTTATCTTTCCCG 2202

384 alSerSerSerAlaSerSerLeuLeuValGlnLeuIleArgGluGlnTrp 400
|||||
2203 TTTCTCTGTCAGCAAGTCTTCTTACTAGTCCAACTTATACGAGAACAGTGG 2252

401 LysGluSerLeuGlyPheAlaIleProIleValGlyLysGluPheAlaLe 417
|||||
2253 AAAGAAAGTTAGGTTTCGCTATCCCTATTGTCGGAAGGAATTTGCTCT 2302

417 uLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGlyT 434
|||||
2303 TCTCCAGCAGACCTATCTTCAGGAGACTTCTCTTTAGCTACAGGAGAT 2352

434 rpPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePheAlaTyr 450
|||||
2353 GGTTGCGAGACTTTGCTGATCCTATGCAATTTCTAACGATCTTTGCTTAT 2402

451 ProSerGlyValProProTyrAlaIleAsnHisLysAspPheLeuGluI1 467
|||||

```

```

/codon_start=1
/transl_table=11
/product="peptide ABC transporter, periplasmic
peptide-binding protein, putative"
/protein_id="AAF38391.1"
/db_xref="GI:7189485"
/translation="MKKISGVGICITILLISLVQVLCQSSSSHSSTSRGELAINRDEP
RSLDPROVRLLSISLVKIHYBGLVQVNNISGNIETALPAEDYSSLDGLTYTFKLKSA
FWSNGDPLAEDTIESKWQAVQVSYTYAFALNPKNVKRIQEGHLSIDREGVHPSNP
ESTLVLPLESTPQSHFKLLALPVFFVPSKQRTLOSQSLPIASGAFVYKPNTKOKOMTK

```

```
complement(4620, .5978)
```

CDS

```

complement(4620..5978)
/gene="CP0577"
/notes="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38395.1"
/db_xref="GI:7189489"
/translations="MKRHSFDSSTKKAIVSKAIQIKIMETDPSLNVETNAEII
ESILQETKEIKLSQAEGLGKYSQETLSNLENTNASIKLSIGSVTEELASLK
OLVDESLEESIQODOLIOSVLBIETSDKFLSSIGETLSGNLDNQNVOGLIKENPE
KSEAASVYVOTLLEPLSKRGETHKKVAHDVNISSLQPHMSVAGSRFRGHDMMG
YRVLGEPKNGEDAVSKDYLERVYSQQLIDKVEDKPTKKNKGLLYSGOTSPKLE
GFLPLGLLISGSGFFWKSASNDGSPFSALRHKETEDSDCFQITSTLISGNQAG
YTKWLSLAVLPSPISFQIEKPEVQLSVLYSYEDWLPIDNIPNKSQPRIPALILGQTM
LAGQKDYILELAHQTNQTLKISPNCRSFRSLQLKOTNQFENSPVDYFIVHAHSCWS
Gf"
complement(6069..9497)
/gene="CP0578"
/notes="conserved hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF38396.1"
/db_xref="GI:7189490"
/translations="MKRRSLKILGICGSSIVLGLPLPOLLSTSGKVLVPSLIH
KESGLSCSAEELKISWFGROTARKIKLTGKADEVPSAEKPELDGSLILILVKKPKG
ITLGSNKLINEPASIDHPVSVDLPGSLITVYLNDCKILISEHGFITMKTVSGSLSVS
GFYLEKSSSEKFKTKVSEDOQSNIPIESLVDSIAQFSSVPVAFKFIASFPF
WDHLLSYEDINLSAEATHNDKISMTAGSEGNQIOMKQACHIKSTFYIVGSSSF
TELKPELASALCNQILIPSTPITSKQIHATVSVAKTIPLDITWKKHIEITSOALPEVA
IHPKDNALQURDKLGIKTEKFSIDRYSSSTVLGGASPSHLNGLISIDNKKHLTK
FRQQAQPLHTVLRALFPOPEVI NVPLDVAVYSIAETGYKNHLEADALDNLPLKL
SCMSGAWKNFLFGKGYTFHNKKQWELPSHPESYAEAFSCKAQIITDNLFPKFSG
KITARENELLHAKGSPNEPIKETPISILHGQCSPLSLVSNHLUPLKLLKLFPS
FNDGSKFTYKNLQALLENPDYDLNNTRILIPDLILSLDESSTSPSKDKIKIQSGG
EIPSLSTITYTKQVRPLPYFGSGDLNFVWYNPQDNKILTLNSFKSALLGGE
LKLVMDFSKLSSTGQGTLOWEVSERYAFKNAKSCPTCLLHRTANRLDLSKLSG
PEETKGLSCLITLAAAGLGSLAETPLIFVQNVSKETPIIINDEKGSRRANNLDAKLEY
DLKQCLAPRODSKTLAEFLSEGOVHLSPSPERFKOTANWHIPIPSFIAGIWMSP
GLKAQISSLAGPRINVSINKNFRGEGPVDIMVDSENLOAQIPLILNEKISILLRENLT
AHLISNEDNKRPLAGAYSOYVTELEIDKQNFYLPPIRYSFEERIQSAT
LDMKISTANTGTWYALFQELDIDQKQFESWFTPIFFSVQKSLICKXKDALIDRR
IRLAWGSLTIAHDLRFLMTLCIDPEVIKKYFNNTSLKTKNFKFLIKIRGSISSPEVDWS
SAVARTALLKYSVLCNPFSSLDKLFSSLDGSDTPPTVHTPFPFKSFDSEIK"
complement(9494..10786)
/gene="CP0579"
/notes="conserved hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF38397.1"
/db_xref="GI:7189491"
/translations="MFRKLPFSKTKTKOKRLNRNGLLOAIQISKVLHNEAKEA
CVLSYVGLLTVCVPLVPLFLRSLQHLNLNKEWLLIKFPDYKKPIVAIVEAAYHATE
SNIGLVVGSFVFCWAGILMLLSLEDGLNKIFRISWTPPISLKRKLVSFYVTLVSPMI
FIIVGSMYITQILQIYOKAFSLSHSMALYFISREVPYLLIYALPCFCYAFIPLRV
AQKQSLIITLIGSVYIVFOKAFSLQVSIYFNYSFTYGAIVLPSFLILLYITWMI
YLFGGALFTIIONRGCTFIPLGKILPSCYLIQITSTYVILATITROPNEGLSPIAQF
IAQSKVPVIGVSOCLDLEKGFLEFPYNNGYQVPEFSELTINDIADKLLHREIFKK
FNPDLGITFIENSQNTFNQASKNKENUTLSEIARK"
complement(10849..11649)
/gene="CP0580"
/notes="conserved hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF38398.1"
/db_xref="GI:7189492"
/translations="MFRKLPFSKTKTKOKRLNRNGLLOAIQISKVLHNEAKEA
CVLSYVGLLTVCVPLVPLFLRSLQHLNLNKEWLLIKFPDYKKPIVAIVEAAYHATE
SNIGLVVGSFVFCWAGILMLLSLEDGLNKIFRISWTPPISLKRKLVSFYVTLVSPMI
FIIVGSMYITQILQIYOKAFSLSHSMALYFISREVPYLLIYALPCFCYAFIPLRV
AQKQSLIITLIGSVYIVFOKAFSLQVSIYFNYSFTYGAIVLPSFLILLYITWMI
YLFGGALFTIIONRGCTFIPLGKILPSCYLIQITSTYVILATITROPNEGLSPIAQF
IAQSKVPVIGVSOCLDLEKGFLEFPYNNGYQVPEFSELTINDIADKLLHREIFKK
FNPDLGITFIENSQNTFNQASKNKENUTLSEIARK"
complement(10849..11649)
/gene="CP0580"
/notes="conserved hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF38399.1"
/db_xref="GI:7189493"
/translations="MFRKLPFSKTKTKOKRLNRNGLLOAIQISKVLHNEAKEA
CVLSYVGLLTVCVPLVPLFLRSLQHLNLNKEWLLIKFPDYKKPIVAIVEAAYHATE
SNIGLVVGSFVFCWAGILMLLSLEDGLNKIFRISWTPPISLKRKLVSFYVTLVSPMI
FIIVGSMYITQILQIYOKAFSLSHSMALYFISREVPYLLIYALPCFCYAFIPLRV
AQKQSLIITLIGSVYIVFOKAFSLQVSIYFNYSFTYGAIVLPSFLILLYITWMI
YLFGGALFTIIONRGCTFIPLGKILPSCYLIQITSTYVILATITROPNEGLSPIAQF
IAQSKVPVIGVSOCLDLEKGFLEFPYNNGYQVPEFSELTINDIADKLLHREIFKK
FNPDLGITFIENSQNTFNQASKNKENUTLSEIARK"

```

gene

CDS

Glimmer2; putative\*

/codon\_start=1

alignment\_scores:

Quality: 2739.00 Length: 532  
 Ratio: 5.148 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

```

US-09-824-567-2 x AE002216/rev ..
Align seg 1/1 to reverse of: AE002216 from: 1 to: 11764
1 MetArgLysIleSerValGlyIleCysIleThrIleLeuLeuSerLeuSe 17
1707 ATGCGCAAGATATCAGTGGGAATCTGTATCACCATTCTCTAGCCTCTC 1658
17 rValValLeuGlnGlyCysLysGluSerSerHisSerSerThrSerArg 34
1657 CGTAGCTCTCCAGGCTGCAAGGAGTCCAGTCACTCCTTACATCTCGGG 1608
34 LyGluLeuAlaIleAsnIleArgAspGluProArgSerLeuAspProArg 50
1607 GGAAGACTCGCTATTATATAGAGATGAACCCCGTCTTTAGATCCAGA 1558
51 GlnValArgLeuLeuSerGluIleSerLeuValLysHisIleTyrGluG 67
1557 CAAGTGGGACTCTTTTCAGAAATAGCCCTGTCAACATATATATATAGG 1508
67 yLeuValGlnGluAsnAsnLeuSerGlyAsnIleGluProAlaLeuAla 84
1507 ATTAGTTCAAGAAATAATCTTTTCAGAAATATAGAGCTTCCTTTGCG 1458
84 LuAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLysLeuLys 100
1457 AAGACTACTCTCTTCTCGGACGACTCACTATATCTTTTAAACTGAAA 1408
101 SerAlaPheTrpSerAsnGlyAspProLeuThrAlaGluAspPheIleG 117
1407 TCAGCTTTTGGAGTAAATGGCAGCCCTTACAGCTGAAGACTTATAGA 1358
117 uSerTrpLysGlnValAlaThrGlnGluValSerGlyIleTyrAlaPheA 134
1357 ATCTTGAAACAAGTAGCTACTCAAGAAGTCTCAGGAATCTATGCTTTG 1308
134 LaLeuAsnProIleLysAsnValArgLysIleGlnGluGlyHisLeuSer 150
1307 CCTTGAATCCAAATTAATAATGTAAGAAAGATCCAAAGAGGACACCTCT 1258
151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValValTh 167
1257 ATAGCACTTTGGAGTGCCTCTCTTAATGAATCTACACTTTGTTTAC 1208
167 rLeuGluSerProThrSerHisPheLeuLysLeuLeuAlaLeuProValP 184
1207 CTTGGAATCCCAACCTCGCATTTCTTAAAACTTTAGCTCTTCCAGTCT 1158
184 hePheProValHisLysSerGlnArgThrLeuGlnSerLysSerLeuPro 200
1157 TTTTCCCGCTTCAATAATCTCAAGAAACCTCGCAATCCAAATCTTACCT 1108
201 IleAlaSerGlyAlaPheTyrProLysAsnIleLysGlnLysGlnTrpIl 217
1107 ATAGCAAGCGGAGCTTCTATCTCTTAAATAATATCAAAACAAATGAT 1058
217 eLysLeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValGluThrL 234
1057 AAAACTCTCAAAAAACCTCTACTACTATATCAAGTCAAGTGGGAAC 1008
234 yThrIleThrIleHisPheIleProAspAlaAsnThrAlaLysLeu 250
1007 AAACGATTACGATTCACTTCTTATCCCGATGCAACACACAGCAGCAAA 958

```

```

251 PheAsnGlnGlyLysLeuAsnTrpGlnGlyProProTrpGlyGluArg11 267
|||||
957 TTTAATCAGGGAACATCAATTGGCAAGACCTCTTGGGGAAGCGCAT 908
|||||
267 eProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerP 284
|||||
907 TCTTCAAGAAACCTATFCCAATTTACAGTCTAAGGGCACCTTACACTCTT 858
|||||
284 heAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhePro 300
|||||
857 TTGATGTCGCGAGNACCTCATGGCTACCTTCAATATCAATAAATTCGCC 808
|||||
301 LeuAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysG1 317
|||||
807 CTCACCAATATGAAGCTTAGAAGACCTTAGCATCAGCTTAGATAAGA 758
|||||
317 uAlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisL 334
|||||
757 AGCTCTTGCTCAACTATATCTTAGCGCGTGCAAAACCTGGCGATCATC 708
|||||
334 euLeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMet 350
|||||
707 TCTTACCTACAAAATATTCATAGCTATCCGGAACATCAAAAACAGAGATG 658
|||||
351 AlaGlnArgGlnAlaThrAlaLysLysLeuPheLysGluAlaLeuGluG1 367
|||||
657 GCACACGCCCAAGCTTACGCTTAAAAAACTCTTTAAAGAGCTTTTAAAGA 608
|||||
367 uLeuGlnIleThrAlaLysAspLeuGluHisLeuAsnLeuIlePheProV 384
|||||
607 ACTCCAAATCAGCGCTAAGATCTCGAACATCTTAATCTTACTTCTCCG 558
|||||
384 alSerSerAlaSerSerLeuValGlnLeuIleArgGluGlnTrp 400
|||||
557 TTTCTCTCGTCAGCAAGTCTTTACTAGTCCAACTTATACGAGACAGTGG 508
|||||
401 LysGluSerLeuGlyPheAlaIleProIleValGlyGluPheAlaLe 417
|||||
507 AAAGAAAGTTTAGGTTCGCTATCCCTATGTCGGAAGGAATTTGCTCT 458
|||||
417 uLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGlyT 434
|||||
457 TCTCAAGCAGACCTATCTTCAGGGAACCTCTCTTTAGCTACAGAGGAT 408
|||||
434 rpPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePheAlaTyr 450
|||||
407 GGITCGGACACTTGTGTCATCTATGCACTTCTTAACGATCTTTGCTTAT 358
|||||
451 ProSerGlyValProProTyrAlaIleAsnHisLysAspPheLeuGluI1 467
|||||
357 CCATCAGGAGTTCCTCCTTATGCAATCAACCAATAGGACTTCTTAGAAT 308
|||||
467 eLeuGlnAsnIleGluGlnGluAspHisGlnLysArgSerGluLeuV 484
|||||
307 TCTACAAAACATAGAACAGCAGATCACCAAAAACGCTCGGAATTAG 258
|||||
484 alSerGlnAlaSerLeuTyrLeuGluThrPheHisIleIleGluProIle 500
|||||
257 TGTCCGACGCTCTCTTTACCTAGACCTTTTCATATTATTGAGCGGATC 208
|||||
501 TyrHisAspAlaPheGlnPheAlaMetAsnLysLysLeuSerAsnLeuG1 517
|||||
207 TACCAGCGCATTTCAATTTGCTATGTAATAAAAACCTTCTTAATCTAGG 158
|||||
517 yValSerProThrGlyValValAspPheArgTyrAlaLysGluAsn 532
|||||
157 AGTCTCACCACAGAGTTGTGGACTTCGGTATCTATCTAAGGAAAT 112
|||||

```

seq\_name: gb\_ba:AP002545

seq\_documentation\_block:

LOCUS AP002545

DEFINITION Chlamydothila pneumoniae J138 genomic DNA, linear BCT 08-DEC-2000

## ACCESSION

AP002545 AB033780 AB033781 AB033792 AB033793 AB033794 AB033795  
 AB033796 AB033797 AB033798 AB033799 AB036071 AB036072 AB036073  
 AB036074 AB036075 AB036076 AB036077 AB036078 AB036345 AB036346  
 AB036347 BA000008  
 AP002545.2 GI:9956082

## KEYWORDS

Chlamydothila pneumoniae J138 (strain:J138) DNA.  
 Chlamydothila pneumoniae J138

## ORGANISM

Bacteria; Chlamydiales; Chlamydiaceae; Chlamydothila.

## REFERENCE

1 (sites)

## AUTHORS

Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K.,  
 Shiba, T., Ishii, K., Hattori, M., Kuhara, S. and Nakazawa, T.

## TITLE

Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 from Japan and CWL029 from USA

## JOURNAL

Nucleic Acids Res. 28 (12), 2311-2314 (2000)

## MEDLINE

20330349

## REFERENCE

2 (bases 1 to 299650)

## AUTHORS

Shirai, M.

## TITLE

Direct Submission

## JOURNAL

Submitted (04-JUL-2000) Mutsunori Shirai, Yamaguchi University  
 School of Medicine, Department of Microbiology; 1-1-1  
 Minamikogushi, Ube, Yamaguchi 755-8505, Japan  
 (E-mail: mshirai@po.cc.yamaguchi-u.ac.jp, Tel: 81-836-22-2227,  
 Fax: 81-836-22-2415)

## COMMENT

On Aug 31, 2000 this sequence version replaced gi:6172286  
 gi:6172288 gi:6172310 gi:6172312 gi:6172314 gi:6172316 gi:6172318  
 gi:6172320 gi:6172322 gi:6172324 gi:6635158 gi:6635160 gi:6635162  
 gi:6635164 gi:6635166 gi:6635168 gi:6635170 gi:6635172 gi:8547422  
 gi:8547426 gi:8547430 gi:8978373  
 AB033780-AB033781, AB033792-AB033799: Submitted (25-Oct-1999)  
 AB033845-AB038347: Submitted (14-Feb-2000)  
 AB036071-AB036078: Submitted (18-Dec-2000).

## FEATURES

## source

1..299650  
 /organism="Chlamydothila pneumoniae J138"  
 /strain="J138"  
 /db\_xref="taxon:138677"  
 /note="synonym: Chlamydia pneumoniae (strain J138)"  
 complement(1..282)  
 /gene="CP10001"  
 complement(1..282)  
 /gene="CP10001"  
 /codon\_start=1  
 /transl\_table=11  
 /product="CT001 hypothetical protein"  
 /protein\_id="BAA98211.1"  
 /db\_xref="GI:8978374"  
 /translation="MLGKTIIRGLSLIVLCALNVGLIGITHNKLNIATKLCGGVSTP  
 ATQITYIIIGIAGVICLLSFCPCFKSKSRHSHGSDSCSSGCHSHSDKN"  
 573..878  
 /gene="gatC"  
 573..878  
 /gene="gatC"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Glu-tRNA Gln amidotransferase (C subunit)"  
 /protein\_id="BAA98212.1"  
 /db\_xref="GI:8978375"  
 /translation="MEQFHLDRREITLLIAKASALQSEELIOEYOTSLSAVITSNKEA  
 LAIEDADSCSESLFMHVNVNEDLREDSVTSDFNREEFLRNVPESGLGVKVPATK"  
 895..2373  
 /gene="gata"  
 895..2373  
 /gene="gata"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Glu-tRNA Gln amidotransferase (A subunit)"  
 /protein\_id="BAA98213.1"  
 /db\_xref="GI:8978376"  
 /translation="MYRYSALELAKAVTLGELTATGVTQHFFHRIIEAEGQVCAFTSL  
 CKQALAEQLIDKKRSRGEPLGAGVPVGIKDNHTVGLKTKCASRVLENQPPFD  
 ATVERIKKEDGIIILGLKMDDEFANGSTTLYISAFHTHPWDLRSVPGGSSGSSAAAY

## gene

## CDS



233905 GAGAACTCGCTATTATAATAGAGATGAACCCCGTTCTTTAGATCAAGA 233954  
 51 GlnValArgLeuLeuSerGluIleSerLeuValLysHisIleIleYrGluCl 67  
 233955 CAGTGGGAGACTCTTTCAAGAAATAGCCTTGTCAACATATCTATGAGGG 234004  
 67 yLeuValGlnGluAsnLeuSerGlyAsnIleGluProAlaLeuAlaG 84  
 234005 ATTAGTTCAAGAAATAATCTTTCAAGAAATATAGAGCGTCTCTTGAG 234054  
 84 uAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLysLeuLys 100  
 234055 AAGACTACTCTCTTCTCGGAGGACTCACTTATATCTTTTAAACTGAAA 234104  
 101 SerAlaPheTrpSerAsnGlyAspProLeuThrAlaGluAspPheIleG 117  
 234105 TCAGCTTTTGGAGTAAATGGGACCCCTTAACAGCTGAAGACTTTATAGA 234154  
 117 uSerTrpLysGlnValAlaThrGlnGluValSerGlyIleTyrAlaPheA 134  
 234155 ATCTTGAAACAAAGTAGCTACTCAAGAGTCTCAGGAATCTATGCTTTTG 234204  
 134 laLeuAsnProIleLysAsnValArgLysIleGlnGluLysHisLeuSer 150  
 234205 CCTGGAATCCAAATAATGTCAAGAGATCCAAGAGGACACCTCTCC 234254  
 151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValValTh 167  
 234255 ATAGACCAATTTGGAGTGCATCTCTTAATGAATCTACACTTGTGTTCAC 234304  
 167 rLeuGluSerProThrSerHisPheLeuLysLeuLeuAlaLeuProValP 184  
 234305 CCTGGAATCCCAACTCGCATTTCTTAAAACTTTTAGCTCTCCAGTCT 234354  
 184 hePheProValHisLysSerGlnArgThrLeuGlnSerLysSerLeuPro 200  
 234355 TTTTCCCTCCCTTCAATAATCTCAAGAACCCCTGCAATCCAAATCTCTACT 234404  
 201 IleAlaSerGlyAlaPheTyrProLysAsnIleLysGlnLysGlnTrpIl 217  
 234405 ATAGCAAGCGGAGCTTCTATCTCAAAATAATCAAAACAAAACAATGGAT 234454  
 217 eLysLeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValGluThrL 234  
 234455 AAAACTCTCAAAACCCCTCACTACTATATCAAAAGTCAGTGGAAACTA 234504  
 234 yThrIleThrIleHisPheIleProAspAlaAsnThrAlaAlaLysLeu 250  
 234505 AAACGATTACGATTCACTTCATTCCTCGATGCAACACAGCAGCAAAACTA 234554  
 251 PheAsnGlnGlyLysLeuAsnTrpGlnGlyProProTrpGlyGluArgIl 267  
 234555 TTTAATCAGGGANAACTCAATTGGCAGGACCTCTCTTGGGAGAACGCAT 234604  
 267 eProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerP 284  
 234605 TCCTCAAGAACCCCTATCCAAATTACAGTCTAAGGGGCACCTATACACTCT 234654  
 284 heAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhePro 300  
 234655 TTGATGTCCGAGGAACCTCATGCTCACTTCAATATCAATAAATCCCC 234704  
 301 LeuAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGl 317  
 234705 CTCACAATATGAAGCTTAGAGAGCCCTTAGCATCAGCTTAGATAGGA 234754  
 317 uAlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisL 334  
 234755 AGCTCTGTCTCACTATATCTTAGGCGTGCAAAACTGCGGATCATC 234804  
 334 euLeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMet 350  
 234805 TCCTACCTACAAATATTCATAGTATCCCGACACATCAAAACACAGAGATG 234854

351 AlaGlnArgGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuGluG 367  
 234855 GCACACGCCAAGCTTACGCTAAAAAATCTTTTAAAGAAGCTTTAGAGA 234904  
 367 uLeuGlnIleThrAlaLysAspLeuGluHisLeuAsnLeuIlePheProV 384  
 234905 ACTCCAAATACCCCTAAGATCTCGAACATCTTAATCTTATCTTTCCCG 234954  
 384 alSerSerAlaSerSerLeuValGlnLeuLeuIleArgGluGlnTrp 400  
 234955 TTTCTCTCGTCAGCAAGTCTTTACTAGTCCACTTATACGAGAACAGCTGG 235004  
 401 LysGluSerLeuGlyPheAlaIleProlleValGlyLysGluPheAlaLe 417  
 235005 AAAGAAATTTTAGGGTTTCGCTATCCCTATTTCGGAAGGAATTTGCTCT 235054  
 417 uLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyLyt 434  
 235055 TCTCCAGCAGACCTATCTTCAGGGAACTTCTTTAGCTACAGGAGGAT 235104  
 434 rpPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePheAlaTyr 450  
 235105 GGTTCGAGACTTTGCTGATCTCTATGGCATTTCTAACGATCTTTGCTTAT 235154  
 451 ProSerGlyValProProTyrAlaIleAsnHisLysAspPheLeuGluI 467  
 235155 CCATCAGAGTTCCTCTTATGCAATCAACCATAGGACTTCTTAGAAAT 235204  
 467 eLeuGlnAsnIleGlnGlnGlnAspHisGlnLysArgSerGluLeuV 484  
 235205 TCTCAAAACATAGACAGACAGACAGATCACCANAACGCTCGAATTAG 235254  
 484 alSerGlnAlaSerLeuTyrLeuGluThrPheHisIleIleGluProIle 500  
 235255 TGTGCGAAGCTTCTCTTTACCTAGAGACCTTTCATATTATGAGCGGATC 235304  
 501 TyrHisAspAlaPheGlnPheAlaMetAsnLysLysLeuSerAsnLeuG 517  
 235305 TACCACACGCATTTTCATTTGCTATGAATAAAAAAATTTCTTAATCTAGG 235354  
 517 yValSerProThrGlyValValAspPheArgTyrAlaLysGluAsn 532  
 235355 AGTCTACCAACAGAGAGTTGTGACTTCCGTTAAGTAAGGAAAAAT 235400  
 seq\_name: gb\_ba:AE001293  
 seq\_documentation\_block:  
 LOCUS AE001293 11944 bp DNA linear BCT 30-OCT-2000  
 DEFINITION Chlamydia trachomatis section 20 of 87 of the complete genome.  
 ACCESSION AE001293 AE001273  
 VERSION AE001293.1 GI:3328597  
 KEYWORDS  
 SOURCE Chlamydia trachomatis.  
 ORGANISM Chlamydia trachomatis  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 REFERENCE 1 (bases 1 to 11944)  
 AUTHORS Stephens,R.S., Kaiman,S., Lammel,C.J., Fan,J., Marathe,R.,  
 Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,  
 Koonin,E.V. and Davis,R.W.  
 Genome sequence of an obligate intracellular pathogen of humans:  
 Chlamydia trachomatis  
 Science 282 (5389), 754-759 (1998)  
 JOURNAL 99000809  
 MEDLINE 9784136  
 PUBLISHED  
 REFERENCE 2 (bases 1 to 11944)  
 AUTHORS Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,  
 Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.  
 Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
 Nat. Genet. 21 (4), 385-389 (1999)  
 JOURNAL 99206606  
 MEDLINE 10192388  
 PUBLISHED  
 REFERENCE 3 (bases 1 to 11944)



## AUTHORS

Stephens, R. S., Kalman, S., Lammel, C. J., Fan, J., Marathe, R.,  
 Aravind, L., Mitchell, W. P., Olinger, L., Tatusov, R. L., Zhao, Q.,  
 Rooin, E. V. and Davis, R. W.  
 Direct Submission  
 Submitted (20-MAY-1998) Program in Infectious Diseases, University  
 of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA

## FEATURES

Location/Qualifiers

1..11944

/organism="Chlamydia trachomatis"

/strain="D/UW-3/CX"

/db\_xref="taxon:813"

complement(105..878)

/gene="CT192"

complement(105..878)

/gene="CT192"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAC67784.1"

/db\_xref="GI:3328598"

/translation="MFFSGSVETLPSTCLISKKTKKEEFVRSVQGEASQNTLSWTIR  
 PRPSIQGSKVSALFVIGTVLAIAGACAAAGGAFSVGLVFLGVLATGLLL  
 AVLEFCHIRSSMEKYVILTKQDLFKEVIEQATLIEFASYTCPCGIPLSGPEVQ  
 QERVLQDLIDLSHPKYIAGVSHVVELVAKGIGRNGERLLEBGIDTQDNFVRWD  
 ELFTLGQGMVRLDQFCCKVLPKKKSEINNVNDC"

complement(905..2023)

/gene="tgt"

/note="CT193"

complement(905..2023)

/gene="tgt"

/codon\_start=1

/transl\_table=11

/product="Queuine tRNA Ribosyl Transferase"

/protein\_id="AAC67785.1"

/db\_xref="GI:3328599"

/translation="MALRFEILHQSKSRARVGRITAHGYIDTPAFVPVATNALKG  
 VLDSNIPMFCNTYHLVHPGAEATAAGGLHOFGRNAPIITDSGSGFOLFSLAYGS  
 VADEIKSCGKKGGNTIKVNDGQVHKSVROGRKLFLSPSLVQAQKDLGADILPL  
 DELLPHADPTVHOSORTVYWEKSLDYHLKNGIQSMYGVTHGGTFDQRLGCK  
 FVELDFDPSAIGGSLGNLQDIVVVGVTAANLSAERHLLGIGDLPSIWAIVGFG  
 IDSDSYTPAARHGMLITSQGLKNNORVSSDNLPIEFQCSCLACSGGITRAYLR  
 HLFKVEPNAGIWAISHNMHMQVMKREIREGLNDRI"

complement(2150..3562)

/gene="tgtE"

/note="CT194"

complement(2150..3562)

/gene="tgtE"

/codon\_start=1

/transl\_table=11

/product="Agg+ Transporter (CBS Domain)"

/protein\_id="AAC67786.1"

/db\_xref="GI:3328600"

/translation="MDSKTSHLDELCKLEBAFTTLFAGEHSODLTSIVSVYNPIDL  
 AVASCLDSRLIYNKIOSIASKAFIINTDSASRWAIFRNLSDEICALITEQMP  
 DEAWVLDDIPDRYRILDLIDKALKIRDLQKHGNTAGRLMTNEFFALMETTV  
 KEVATCINNEGIDLTLFVDFKGLQFVDSRLIIASPEMLKQIMRPIEHLV  
 ADTTEEVVDLVERKYAVLPVWDEENLIGALTYEDVETIEDIADETIARMAGTE  
 DVGHDHCVORFTLRAPWLLITLCAGLYSASVAYOKAPILLAAVFFIPLVNL  
 SGNVGVQCSLTVRSMTATGLSGRRRETILKEMSIGLLTGVALGICGLVWCOMGL  
 GLGLPATGVQGVTVSVGLIGASLTATTLGLVSPFTFAKTGVDPALASGPVITLND  
 INSMWIFLLITGLNLVLFK"

complement(4001..5092)

/gene="CT195"

complement(4001..5092)

/gene="CT195"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAC67787.1"

/db\_xref="GI:3328601"

/translation="MWSMSLNLPPAEVLRPVTAKSPFNSAWRATQKRTGCSLNAS  
 STRPSSARVSVFVGLVWLVGICFRIAYLANSRVLTFPRKFAIILPIFLPGGLAI  
 LFRTAGKVDLYGKKIQFTSRQWVERVILCEKEGEFIRPIQEPDMYMDYSLLDRTGSG

gene

CDS

IAPVVTPMDARTVICWTSILLPIFSIVRMLYNIPREFIVPFYILFQWVRNXYOTD  
 IKFERFVCSDIVMERISLQAVKAPFYGAVCYLANLYGLINPLISGRVVLASLEWD  
 NNDVRSFGVWGIECEKNYFEGGTRSGLGQNAWYLLGCPQVQLFLLADGVIIISGA  
 RSIQSPSESEYLAFLYGAVPRLAGF"

5316..5636

/gene="CT196"

5316..5636

/gene="CT196"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAC67788.1"

/db\_xref="GI:3328602"

/translation="MATTVNPNYSILPFCEKMWSSATLAAHFFNHKILILPLLGVFC  
 ILLGALLITGVITIFSIASVFLSLGVTLVIVAGLCAAFKAPLFSVTSKASTLLH  
 TSCD"

5712..6728

/gene="gcp\_1"

/note="CT197"

5712..6728

/gene="gcp\_1"

/codon\_start=1

/transl\_table=11

/product="O-Sialoglycoprotein Endopeptidase"

/protein\_id="AAC67789.1"

/db\_xref="GI:3328603"

/translation="MLTGLGSSCDTSCSLVQNGKTLANKIASODIHASYGVPEL  
 ASRAHLQTFPELLTAATQSAGVSLIEDILISYVANTPGLIGALSIGNFRAGLASLAR  
 PLIGVNHVEAHLAACMEAPATQFALGLAISAHTSLFLMPDATTLFLIGLRDADL  
 GETDKVARELGLPYGGOKLEELADEGDAFAFSPARYSGYDFSGLKATVIAL  
 KGNSSAKAPPEVSETOKRNIAASFQKAVMTIAQKLPDIVKTFCESLIVGGGVAN  
 NSYFRLLNQICSLPIYFPFSQCSDNAAMIAGLRGLFCNRTHVSKEVIPCARYQWE  
 SACS"

6692..8248

/gene="oppA\_3"

/note="CT198"

6692..8248

/gene="oppA\_3"

/codon\_start=1

/transl\_table=11

/product="Oligopeptide Binding Protein"

/protein\_id="AAC67790.1"

/db\_xref="GI:3328604"

/translation="MKKISVGICILLALATSGCKSSSNATHRSRPATHVAVSVKDDP  
 WFDPREVRLLSDINLIHLVGLVQETPSGEVPEALAESFSLSEDKTYTFNLKAF  
 RSGDLITAHDFVRWNWDLONRVASLYSFAFLPIDWNKDSGFPKADHTVITNLLTP  
 TPEFLKLLTLPVYPVHSQHQIRKEKSLPSTGAFFLAEKKDRWLKLEKSPYYNK  
 DQVAVQICIIIPDQQTASALFNOGLDWGGLPWGHSIPQETLATTNKRPRSFDI  
 SGTSWLFNTAKKPFSSKRLQALSLVNLKEALASLAFYKPAKHLPAHLHTYPEQS  
 YKQOEAITLAKSLLEALTLELMTIEDLEKPLTFSTATMSQIAQMLRQWRSLG  
 ITFPGCKEYALLQNDLIGNTFMSIGWFAFSDPLAFSLFSSKGVKPVALQDPF  
 DQLILSTETKPNQKRSALISEASLYIERQNVIEPLYHDFVHTYTNKLSFVRLHPSG  
 LVDMRYAKNS"

8407..9348

/gene="oppB\_1"

/note="CT199"

8407..9348

/gene="oppB\_1"

/codon\_start=1

/transl\_table=11

/product="Oligopeptide Permease"

/protein\_id="AAC67791.1"

/db\_xref="GI:3328605"

/translation="MLSVIKRRLFLNLSLWVVVTLTFPIIKTPGDPFNDENGNIIS  
 SETLALLKNRYGLDKPLETOYLYLKLCLTLFDGESLIYKORTVISTIAALPSSAIL  
 GLESLCGLFGGITGLIILAFYKKSQRTIFFSSIQISVFAFVIGAFIYVFAIKYS  
 CLPIACWGNFHTLPLSALAITPMATITQLTCASVSNLKKDYVLAYAKGLSPFKV  
 LKHLPLAFPVISYSAFLITMTQFTSIENLFCIPGLCKWFCISIKQDYPITLG  
 LSVFYCAFMULTSLCCDLQAQWIDPQIRYSYKERSK"

9360..10225

/gene="oppC\_1"

/note="CT200"

9360..10225

gene

CDS

```

/gene="oppC_1"
/codon_start=1
/transl_table=11
/product="Oligopeptide Permease"
/protein_id="AAC67792.1"
/db_xref="GI:3328606"
/translation="MFRSSSHNTWRYIFRNKMLVGLCTIAVLAAAFLLPYLPDYE
RTFPEHALQSPKAPFFGTDNRCMLARTQGLRLSLIAVSATLIDVCLGLLWSTL
ALASGKXVADIMSRIDEILFIPRIPIVILLYFNHGLPLIATITGHPARII
YGFLLLENKEFVLSARALQASTPHILRKHLLPNSLGPISLITIPNAVITFAFIS
ELGIGIOPPVASLCTLVKEGILHSLAYHPWLFFIPSPFMIIVSVFNCTGEGRLTKLE
ENTLV"
gene 10218..11051
/feature="oppd"
/feature="oppd"
/feature="CT201"
10218..11051
/gene="oppd"
/codon_start=1
/transl_table=11
/product="Oligopeptide Transport ATPase"
/protein_id="AAC67793.1"
/db_xref="GI:3328607"
/translation="MSEDLKIDNLVSVKDSNQRVNHLSLTIKRCOSMALVGENS
KXITVSAVLGFLPDNCFYIQSGRILYSTDITLRSRQLQITIRGKKIATIFQNAWGTL
TPSMRVGAQIVETLRRHFDMSKEAFSKRELLESVHIESPRCLQLYFPFLSGMCO
RVSTALATNPFLIADPEPTALDSISQAQVLRVLTQIHONHSTALLIITHNLAVS
ELACEMALIRYGEIVQGPVQVELLHSPSHPYTQOLIRAIKPIKPSYLSPKREPLATTA
Y"
gene 11066..11809

```

## alignment\_scores:

Quality: 1364.00 Length: 537  
 Ratio: 3.187 Gaps: 8  
 Percent Similarity: 79.702 Percent Identity: 50.093

## alignment\_block:

US-09-824-567-2 x AB001293 ..

Align seg 1/1 to: AB001293 from: 1 to: 11944

```

1 MetArgLysIleSerValGlyIleCysIleThrIleLeuLeuSerLeuSe 17
|||||
6692 ATGCGCAAGATATCAGTGGGATCTGCTGCTCTAGCATAGCAACTTC 6741
17 rValValLeuGlnGlyCysLysGluSerSerHisSerThrSerArg 34
|||||
6742 T.....GGATGTTCAAAATCCTCTCTAACGCAACCCATCGT 6779
34 LY.....GluLeuAlaIleAsnIleArgAspGluProArgSer 46
|||||
6780 CTCACGCTACTCACACAGTTCCTGTAAGCGTAAAGATGATCCTCGACA 6829
47 LeuAspProArgGlnValArgLeuSerGluIleSerLeuValLysH 63
|||||
6830 TTGATCCTCGAGAGTTCGGCTTCTTCGATATCAATTTGATTCATCA 6879
63 sIleTyrGluGlyLeuValGlnGluAsnLeuSerGlyAsnIleGlu 80
|||||
6880 TCCTATGATGAAGATTGGTACAGAAACTCCT...TCTGGAGAAGCTTCC 6926
80 roAlaLeuAlaGluAspTyrSerLeuSerSerAspGlyLeuThrTyr 96
|||||
6927 CTCTTTACGGAGAGTTTCTCTTCATCCGAAGATAAAAAAATACTACT 6976
97 PheLysLeuLysSerAlaPheTyrSerAsnGlyAspProLeuThrAla 113
|||||
6977 TTCAACTTGAAAAAGCTTTTGGACCAATGGAGATCTTATACCGCTCA 7026
113 uAspPheIleGluSerTrpLysGlnValAlaThrGlnGluValSer 130
|||||
7027 TGATTTTGTCCTCTCGGATGATGTTTACAAAATCGTCTCGCTAGTA 7076
130 lefYrAlaPheAlaLeuAsnProIleLysAsnValArgLysIleGln 146

```

```

|||||
7077 TTTATCTTTCCTCCTTCTCCCTATT...GAGCTGAATAAG..... 7114
147 GlyHisLeuSerIleAspHisPheGlyValHisSerProAsnGluSerTh 163
|||||
7115 .....GATTCTGATTTTGGCAAGATGATCATAC 7146
163 rLeuValValThrLeuGluSerProThrSerHisPheLeuLysLeuLeu 180
|||||
7147 TCTTGTATCAATCTCTCACTCCAACTCCACATTTCTAAAGCTGCTTA 7196
180 leuLeuProValPhePheProValHis...LysSerGlnArgThrLeuGln 195
|||||
7197 CCTCTCCCGTATTTTATCTGTGTCATCGCAGCATCAGATACGGAAGAA 7246
196 SerLysSerLeuProIleAlaSerGlyAlaPheTyrProLysAsnIle 212
|||||
7247 GAAAAATCTCTTCGATATCTACTCGAGCTTTTCTTAAAGAGAGAA 7296
212 sGlnLysGlnTrpLysLeuSerLysAsnProHisTyrTyrAsnGln 229
|||||
7297 AGACCGAAGATGTTTAAAGCTAGAGAGAGCCCTACTACTATAATAAG 7346
229 erGlnValGluThrLysThrIleThrIleHisPheIleProAspAla 245
|||||
7347 ACCAGTAGCTGACAGGAGATCTGTATACATCATCTCTCTGATCAACA 7396
246 ThrAlaAlaLysLeuPheAsnGlnGlyLysLeuAsnTrpGlnGlyPro 262
|||||
7397 ACTGCTTCTGCTTTATCAACCAAGGAAGCTAGATTGGCAAGSTCTCCC 7446
262 otrpGlyGluArgIleProGlnGluThrLeuSerAsnLeuGlnSerLys 279
|||||
7447 TTGGGACATAGTATTCGCAAGAACTTTAGCCCAACAACAACAGAC 7496
279 lYHisLeuHisSerPheAspValAlaGlyThrSerTrpLeuThrPhe 295
|||||
7497 GAGCTCCCGCATCTTGTATATCTGCTACTTCTGCTTACATTTTAA 7546
296 lIleAsnLysPheProLeuAsnMetLysLeuArgGluAlaLeuAla 312
|||||
7547 ACTGCCAAAAAGCTTTTATCTTCCAAAGCTTCGCCAAGCTTGTAGCT 7596
312 rAlaLeuAspLysGluAlaLeuValSerThrIlePheLeuGlyArgAla 329
|||||
7597 AGTTTAAACAAAGAAAGCTTGTGCTCTCTGCTTTGTTT.....A 7637
329 ysThrAlaAspHisLeuLeuProThrAsnIleHisSerTyrProGlu 345
|||||
7638 AACCTGCAAAACATCTCTCTGTCACATTTGCACACCTACCCAGAGCAG 7687
346 GlnLysGlnGluMetAlaGlnArgGlnAlaTyrAlaLysLysLeuPhe 362
|||||
7688 CCTTCTTATAAGCAACAAGAGCCATCACTTTAGCTAAATCTTTACTAG 7737
362 sGluAlaLeuGluGluLeuGlnIleThrAlaLysAspLeuGluHisLeu 379
|||||
7738 AGAAGCTCTGACTGAGCTTAACATGACTATTGAGGATCTAGAGAAGTATC 7787
379 snLeuIlePheProValSerSerSerAlaSerSerLeuLeuValGln 395
|||||
7788 CTCTTACCTTTCCCGACAGCTCTACTATGAATCAGATCAGATCAGATG 7837
396 lIleArgGlnGlnTrpLysGluSerLeuGlyPheAlaIleProIleVal 412
|||||
7838 TTGCGGATCAGTGGCGAAGAGTTTAGGAATTTACTTTCCCTATCTCGG 7887
412 yLysGluPheAlaLeuLeuGlnAlaAspLeuSerSerGlyAsnPheSer 429
|||||
7888 GAAAGATATGCTTTGTTGCAAAACGATCTTAATAGCAATACTTTCTTTA 7937
429 euAlaThrGlyGlyTrpPheAlaAspPheAlaAspProMetAlaPhe 445
|||||

```

```

7938 TGTCTATAGTGGCGTGGTTCGCCGACTTTCTGACCCCTTAGCGTTTCT 7987
446 ThrIlePheLaTyRProSerGlyValProProTyRAlaIleAsnHisly 462
:::|||||::: :|||::: |||||:::
7988 TCCATTTCCTCC...TCGAAAGGAGTCAAACTTATGCTTACAGATCC 8034
462 SASpPheLeuGluIleLeuGluAsnIleGluGlnGluAspHisGlnL 479
:::|||| :|::: ||||| |||||:::
8035 TCAATTTCATCACTGCTCTCTATAGAACGGGAAACCCCTCAAA 8084
479 ysArgSerGluValSerGlnAlaSerLeuTyRLeuGluThrPheHis 495
||||| |||||::: |||||::: |||||:::
8085 AACGCTCAGCTTAATTCGGAAGCTTCTCTATACATAGAACACAAAC 8134
496 IleIleGluProLleTyRHisAspAlaPheGlnPheAlaMetAsnLysly 512
|||||::: |||||::: |||||::: |||||:::
8135 GTCATAGACCCCTCATACGACGCTGTCATTATATACAAATAATAA 8184
512 sLeuSerAsnLeuGlyValSerProThrGlyValValAspPheArgTyR 529
|||||::: |||||::: |||||::: |||||:::
8185 ACTTCTTGTGTAGACTACATCTCTCGGCCTTGATGATGCGGTATG 8234
529 laLysGluAsn 532
|||||:::
8235 CTAAAAAAGTCT 8245

seq_name: gb_ba:AE002315
seq_documentation_block:
LOCUS AE002315 12173 bp DNA linear BCT 26-MAY-2000
DEFINITION Chlamydia muridarum, section 46 of 85 of the complete genome.
ACCESSION AE002315 AE002160
VERSION AE002315.2 GI:8163226
KEYWORDS
SOURCE
ORGANISM
Chlamydia muridarum.
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE
1 (bases 1 to 12173)
AUTHORS
Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Unayam,L.A., Utterback,T.,
Berry,X., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39
Nucleic Acids Res. 28 (6), 1397-1406 (2000)
20150255
10684935
2 (bases 1 to 12173)
Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Unayam,L.A., Utterback,T.,
Berry,X., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Direct Submission
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr. Rockville, MD 20850, USA
On Jun 1, 2000 this sequence version replaced gi:7190506.
FEATURES
Location/Qualifiers
1..12173
/organism="Chlamydia muridarum"
/db_xref="taxon:83560"
/note="synonym: Chlamydia trachomatis MoPn"
complement(97..792)
/gene="TC0454"
complement(97..792)
/gene="TC0464"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF39314.1"

gene
CDS
complement(97..792)
/gene="TC0454"
complement(97..792)
/gene="TC0464"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF39314.1"

```

```

/db_xref="GI:7190507"
/translation="MDSSCALKNLSSCEVSRIPEVIATRIKEVALVILVLTVAIAG
ACAVAVISVWSIGLTLVVGVLATGELFALMEICRAQYGRQKQIQOQEVLESS
PIREATLPTFIEQAVYSHEPELPLSQEVRCHRPILLQKSGPAETPKFIAVGNVY
ELVKVMIGPMGRGNVPQGLTVRLWDELFAIGRMGELVRLDGFCKKVLPTLGDTS
VSKTSADRASSED"
complement(1227..2345)
/gene="TC0465"
complement(1227..2345)
/gene="TC0465"
/note="similar to GB:AL009126; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="guanine tRNA-ribosyltransferase"
/protein_id="AAF39315.1"
/db_xref="GI:7190508"
/translation="MALRFEILHQSKKSRARVRIETHEGCDIPTAFVPAVATNGALKG
VLDSNPLMFCNTYHLIVHGAEEAAMGGLHQFIQGNAPIITDSGGFQIFSLAYGS
VAEIKSGKKKGKNSIIRKINDEGVNFKSYRGRKFLSPEVSQAQKHLGADIILPL
DELLPFDTPAYFOQSORTYAWERKSLDYHLANGPYOSMYVIHGGTTPDQKLGCC
FVEDLPDGSAGISGLKRLRDIVGVVDVTTANLSIERPHLLIGIDLSWATVGFGC
IDSPSSYPTKAARHGMLTISOGSLKINNQRYSADLAPIEPCACPACSQGISRAYLR
HLFVHEPNAGIOWASIHNMHYMKIMSKIREKILNDOL"
complement(2467..3939)
/gene="TC0466"
complement(2467..3939)
/gene="TC0466"
/note="magnesium transporter; identified by match to TIGR
protein family HMM TIGR00400"
/codon_start=1
/transl_table=11
/product="magnesium transporter"
/protein_id="AAF73559.1"
/db_xref="GI:8163227"
/translation="MGEQRNEEKLDVALTSGNLMDSKTSKSHLDELSPKLEAFDTLT
AGEHQDLSIVSAYNPIDLAYVAVSCLPSPDSRSILYKSLDSTAKIAFIINTDSARW
AIFRLSDGEICALIEQMPPEDEAIWVLDIDIPDRYRILDLIDVKKALKIRDLKHGR
NATRLMTNEFFAFMTVTKVATCIRNPGVDLTRLVFDLDFKGELOGFVTRSLI
IASPEMLKOIMRPEVHKVLADTREFVVDLVERVKAVLPVVDENEFELGATYEDV
VETTEDIADSTIARMAGTTEDGYHCHVQVRFILRAPMLITLACGLSVASVWAYFQ
KIATPLAWIFIPILVNLGSLNGVQCSTILVRSMATGTLSPGRRRETILKMSIGL
LTVAGLIGLGVVCCMGCLGLSTGTGVLGVTVSVGILGASLTATTLGLVSPFFP
AKIGVDPALASGPITVANDINSWVIFLLITGLNLVLFK"
4208..4366
/gene="TC0467"
4208..4366
/gene="TC0467"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF39316.1"
/db_xref="GI:7190509"
/translation="MDKARTVFFPQLQYDEFTVDLOAKSIGLLKTVAVYAKLAAS
QPSQSP"
complement(4328..5410)
/gene="TC0468"
complement(4328..5410)
/gene="TC0468"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF39317.1"
/db_xref="GI:7190510"
/translation="MFLNLPSEAVRIPVTTSEFPBASSKATOKRITSCITLNPYK
TSLTRVCFVGLIIVLGAICCRVAILCENAGLLFSFKILALDSVLGGIGVLCK
IARKIDVLCERIQPFAIRKWEQVILCEKRGQITRIPQDDMYMDVSLLDKQSGTAP
VYTPPAKARTAFICIAILLITVVRVLYNAFLIPFIVFQIRIQLYQEDLPF
EQQFICSDIFRMSRSFQVAKAPYGVACYLASLYLGLLPLSLRIVMASVERDWNND

```

```

VIRSRINGIFCEKNCLLEGGTRSGLGQAHAWYLLGCFQFRLFLKDKGEIYSGARPS
5581. 5946 /gene="TC0473"
IQAFPEKNEYLTSLYGAALGRITSR"
5581. 5946 /gene="TC0469"
5581. 5946 /gene="TC0469"
5581. 5946 /note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF39318.1"
/db_xref="GI:7190511"
/translation="MNYLFLNNQKAKRMOLTIVNTNSRIFPFSNERILPGSTLSTD
SFYSYKILPLLAGSICLLGALILAGTITVTPAIAASFLSLGVILVAGIGLCSV
RPPLHLSINQSDHLIYY"
6012. 7034 /gene="TC0470"
6012. 7034 /note="O-sialoglycoprotein endopeptidase; identified by
match to PFAM protein family HM PF00814"
/codon_start=1
/transl_table=11
/product="O-sialoglycoprotein endopeptidase"
/protein_id="AAF73560.1"
/db_xref="GI:8163228"
/translation="MLTGLESSDETSALVENKGLIANRIASQDIHAAYGGVPIPEL
ASRAHLQIFPKLLAAVAQDAEVSLEDVELISVANTPGLIGALSGVNFPAKGLASGLKK
TLGVNVEAHLYACLESPIRFPALGLAISGHTSLFMPNATFLLIGLSDKDAI
GETDGVARFLGLPFGQKLEELAQGDDEAFYFSAKVSQNDFFSGLKTAIVLYAL
KGNSSAKAPFPEVSETQKRNIAAFQKAAPMTIAQKLPDIIVKAFSCSLVVGCVAM
NRYFRLNLNCTSLTPYFSSQLASQNAAMIAGLERLFCNQTYVYSKEVIPCARYQE
SACLSH"
6992. 8554 /gene="TC0471"
6992. 8554 /note="similar to GB:M57689 SP:P24141 GB:X56347 PID:143603
PID:40005; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, periplasmic
peptide-binding protein, putative"
/protein_id="AAF39319.1"
/db_xref="GI:7190312"
/translation="WRKVSVCIGLLIALATAITGCKSSSNKSNHSSNQSQSVSMKD
DPTFDPREVRLSDINLIHLYEGVQETPGEVEPALAESFELSEKTYTFHLKK
ALMSNGDLITAHDFVRKNDVLQNRASIIYSFAFLPIDLSDCCGFFAKDNHTLVNLH
TPTPHFLKLTLPVYPVPHPEQIRNEAKALPISTGAFCLKEKDRRWLKEKNYY
NKQVAIQEIHVIHQDOTASALFKQGLDMQGLPWGHSIPQALATANKRRTPQSF
DISGWSLFTNTSKVPFSFKLROALSVLNKEALASPTFVKPAKHLPLHLTYPEQ
PTYKQEAIVLAKTLQBLADLNMITIKLEKPLIFSATSVNSQMAOMRDQWRI
LGTFPICGKEYALQNDLTNTFMSINCFWADFSDPLAFLSVFSKGIKPYALQDP
LFQDLILSTETEKDPRKRISLSEASLYLEKQNIIEPLYHDFVHYAANKLSFVRLHP
SGLVDMRYAKNS"
8668. 9612 /gene="TC0472"
8668. 9612 /note="similar to SP:P24138 GB:X56347 PID:580897
GB:AL009126; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, permease protein"
/protein_id="AAF39320.1"
/db_xref="GI:7190513"
/translation="WLCYIKKRLLFNLLSWVTVLTFPIITIPGPPFNDENGNVLS
EETLAIANKRGKGLPFTQYLLKCLTLDFGESLIYKDRVTIGITLAFPSSAIL
GLIESCLALFGGTLHGAAPYKGGCTIFFSIIQISVPFAVIGAFQYIFAIKYS
LIPACWDSFHTLPSIALAIAPAFITQLTYSASVSLKDKDYVILLAYAKGLSPKV
LIKHLVPALESHTLPSIALAIAPAFITQLTYSASVSLKDKDYVILLAYAKGLSPKV
LSVFYGAFFMLISLCLDLQAWIDPQIRYSYKSHSQ"

```

```

gene 9618. 10481 /gene="TC0473"
CDS 9618. 10481 /gene="TC0473"
9618. 10481 /note="similar to GB:M57689 SP:P24139 GB:X56347 PID:143606
PID:40007; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, permease protein"
/protein_id="AAF39321.1"

alignment_scores:
Quality: 1345.00 Length: 538
Ratio: 3.135 Gaps: 9
Percent Similarity: 79.740 Percent Identity: 49.442

alignment_block:
US-09-824-567-2 x AE002315 ..

Align seg 1/1 to: AE002315 from: 1 to: 12173

1 MetArgLysIleSerValGlyIleCysIleThrIleLeuLeuSerLeuSe 17
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
6992 AUGCGCAAGGTATCAGTGGGAATCTGC.....TTGCTTATCGGRTAGC 7035

17 rValValLeuGlnGlyCysLysGluSerSer.....His 29
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7036 AACAGCAATACCCTGCTGCCAATCTCTCTCAACAAATCAAACTACT 7085

29 erSerThrSerArgGlyGluLeuAlaIleAsnIleArgAspGluProArg 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7086 CCTCCTCTTAATCAA...AGCGTATCTGTGAGCATGAAGGACGATCTCTGC 7132

46 SerLeuAspProArgGlnValArgLeuLeuSerGluIleSerLeuVal 62
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7133 AGTTTGTACCTCGAGAGGTAGGCTTCTTCGGATATTAAATTGATCA 7182

62 shisIleTyrrGluGlyLeuValGlnGluAsnLeuSerGlyAsnIleG 79
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7183 CCATCTTTATGAGGGACTGTTTCAAGAGACTCCT...TCGGGAGAGGTTT 7229

79 luProAlaLeuAlaGluAspTyrrSerLeuSerSerAspGlyLeuThr 95
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7230 TCCCTGTCATTAGCAGAAAGTTTCTTCTTATCTGAAGATAAGAAACCTAC 7279

96 ThrPheLysLeuLysSerAlaPheTrpSerAsnGlyAspProLeuThrAl 112
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7280 ACTTTCACCTATAAAAAAGCACTCTGGAGTAATGGAGACCTGATTACTGC 7329

112 aGluAspPheIleGluSerTrpLysGlnValAlaThrGlnGluValSerG 129
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7330 TCATGATTTGTACGTTCTTGAAGCATGTTTTCGAAAAATCCGATTGCAA 7379

129 lyIleTyrrAlaPheAlaLeuAsnProIleLysAsnValArgLysIleGln 145
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7380 GTATCTACTCTTTCGCTTTTACCCTTT..... 7408

146 GluGlyHisLeuSerIleAspHisPheGlyValHisSerProAsnGluSe 162
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7409 .....GATTAGTAAGAAT...TGTTGGATTCTTCGCTAAGGACAATCA 7449

162 rThrLeuValValThrLeuGluSerProThrSerHisPheLeuLysLeuL 179
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7450 TACACTGTTATCAATTTGCATACCCCTACTCCACACTTTCTGAAATTAC 7499

179 euAlaLeuProValPhePheProValHis...LysSerGlnArgThrLeu 194
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7500 TTACCTCTCCAGTATTTATCTCTGTGACCCCAACATCAATTCGGAAT 7549

195 GlnSerLysSerLeuProIleAlaSerGlyAlaPheTyrrProLysAsnIl 211
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7550 GAAGCAAAAGCTCTTCCCAATATCTACAGGGGCTTTTGTGTTTAAAGAGAA 7599

```

```
211 eLysGlnLysGluThrPheLeuSerLysAsnProHisTyrTyrAsnG 228
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7600 AAAAGATTCGGCGTGGCTTAAATAGAAAAAATCCCTATTATATAACA 7649
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
228 InSerGlnValGluThrLysThrLeuThrLeuHisPheLeuProAspAla 244
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7650 AAGAGCAGTTCGATCCAGGAATCCACATACATGATTCCTGATCAG 7699
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
245 AsnThrAlaAlaLysLeuPheAsnGlnGlyLysLeuAsnTrpGlnGlyPr 261
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7700 CAGACCGCTTCTGCTTGTTCACAAAGAAATGGATTCGCAAGGACT 7749
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
261 ProTrpGlyGluArgLeuProGlnGluThrLeuSerAsnLeuGlnSerL 278
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7750 TCCTTGGGACACGACATCCCAAGAGCTTAGCAACAGCAATAAAC 7799
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
278 ysGlyHisLeuHisSerPheAspValAlaGlyThrSerTrpLeuThrPhe 294
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7800 GTCGTACTCTCAATCATTTGATATATCGCACCTCGTGGCTCAGTTC 7849
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
295 AsnLeuAsnLysPheProLeuAsnMetLysLeuArgGluAlaLeuAl 311
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7850 AACACATCAAAAGTACCTTCAGGCATCCCAAGTTCGACAAAGCTTAAG 7899
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
311 aserLeuAspLysGluAlaLeuValSerThrIlePheLeuGlyArgA 328
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7900 CTTGTCTCTAAATAGAGAGCCTTAGCCTCCCAACCTTGT..... 7942
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
328 lalysThrAlaAspHisLeuLeuProThrAsnIleHisSerTyrProGlu 344
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7943 ..AAGCCGCTAAACCTTACTCTACACACCTGCACACTTATCCAGAG 7990
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
345 HisGlnLysGlnGluValAlaGlnArgGlnAlaTyrAlaLysLysLeuPh 361
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7991 CAACCTACTTACAAACAACAGCGCTCATTTTAGCCAAAACCTCTGT 8040
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
361 eLysGluAlaLeuGluGluLeuGlnIleThrAlaLysAspLeuGluHisL 378
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8041 ACAGAGACATTAAAGATCTTATATGACTATAAAGATCTTGAAAT 8090
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
378 euAsnLeuLeuPheProValSerSerAlaSerLeuLeuValGln 394
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8091 GCCCTTGTATTTTTCAGCGACATCTAGCGTCAACTCACAATGCCCAA 8140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
395 LeuIleArgGlnGlnTrpLysGluSerLeuGlyPheAlaIleProIleVa 411
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8141 ATGATGCGGATCAATGCGGAAGATTTTAGGAATAACATTCCTCTATTG 8190
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
411 lGlyLysGluPheAlaLeuLeuGlnAlaAspLeuSerSerGlyAsnPhes 428
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8191 CGGGAAGAGTATGCTTTATACAAATGATTTGACCCCAACACTTTT 8240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
428 erLeuAlaThrGlyGlyTrpPheAlaAspPheAlaAspProMetAlaPhe 444
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8241 TTATGTCATAAACGGTTGGTTTCAGACTTTTCGATCTTTACCTTC 8290
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
445 LeuThrIlePheAlaTyrProSerGlyValProProTyrAlaIleAsnHi 461
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8291 TTGCTGTGTTTTCT...TCTAAGGAATCAAACTTATGCTTACAAGA 8337
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
461 sLysAspPheLeuGluIleLeuGlnAsnIleGluGlnGlnAspHisG 478
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8338 CCCTCTGTTTGACAGCTATTCTCTATAGAAAGGAAAGATCCTC 8387
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
478 lnllysArgSerGluLeuValSerGlnAlaSerLeuTyrLeuGluThrPhe 494
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8388 GAAAGCGTATTCTGATTTTCGAGCTTCCTATACCTAGAAAAACAA 8437
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
495 HisIleIleGluProIleTyrHisAspAlaPheGlnPheAlaMetAsnLy 511
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8438 AATATCATAGAAGCGCTCTACGATGTTTTCATATGCGCAANTAA 8487
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
511 sLysLeuSerSerLeuGlyValSerProThrGlyValValAspPheArgT 528
```

```
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8488 TAAACTTCTTTTGTAGATTACATCCCTCAGGTCTTCTCGATATGGGT 8537
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
528 YrAlaLysGluAsn 532
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8538 ATGCAAAAACACTCT 8551
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: gb_pat:AX349605
seq_documentation_block:
LOCUS AX349605 1584 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 128 from Patent WO0202606.
ACCESSION AX349605
VERSION AX349605.1 GI:18615409
KEYWORDS
SOURCE Chlamydomophila pneumoniae.
ORGANISM Chlamydomophila pneumoniae.
REFERENCE 1 (sites)
AUTHORS Ratti,G. and Grandi,G.
TITLE Immunisation against Chlamydia pneumoniae
JOURNAL Patent: WO 0202606-A 128 10-JAN-2002;
Chiron S.p.A. (IT)
FEATURES
location/Qualifiers
1..1584
/db_xref="taxon:83558"
BASE COUNT 506 a 378 c 256 g 444 t
ORIGIN
alignment_scores:
Quality: 1160.50 Length: 521
Ratio: 2.923 Gaps: 4
Percent Similarity: 76.200 Percent Identity: 43.186
alignment_block:
US-09-824-567-2 x AX349605 ..
Align seg 1/1 to: AX349605 from: 1 to: 1584
15 SerLeuSerValValLeuGlnGlyCysLysGluSerSerHisSerSerth 31
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31 ACTCTTTGGCTTACTTACCTCAGGATG.....TCCCATCCCAATC 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
31 rSerArgGlyCyluLeuAlaIleAsnIleArgAspGluProArgSerLeuA 48
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
72 CTCATAAGGAATTTTGTGGTAAATATGAGGAAATGCCACGCTCTGG 121
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
48 spProArgGlnValArgLeuLeuSerGluIleSerLeuValLysHisIle 64
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122 ATCCTGGAAAACTCTCTCATTTGAGACCAAACTCTAATGCGTCATCTA 171
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
65 TyrGluGlyLeuValGlnGlnAsnLeuSerGlyAsnIleGluProAl 81
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
172 TATGAAGACTCTCGAAGAACATTCCTCCAAATCGAGAGATTAAACACGC 221
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
81 aLeuAlaGluAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheL 98
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 CCTTCGCAAGAGCTACACCATCTCGAAGACGGGACTCGGTACACATTTA 271
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98 ysLeuLysSerAlaPheTrpSerAsnGlyAspProLeuThrAlaGluasp 114
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
272 AAATCAAAAACATCTTTGGAGTAAAGGAGACCTCTGACAGCTCAAGAC 321
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
115 PheLeuGluSerTrpLysGlnValAlaThrGlnGluValSerGlyIleTy 131
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
322 TTTCTCTCTCTTGGAGGAAATCCCTAAAGAGAGATCGCTCCCGGTATA 371
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
131 rAlaPheAlaLeuAsnProIleLysAsnValArgLysIleGlnGluGlyH 148
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
372 TCTCTATGGCTTTTACCTATCAAAAATGCTCGGGCAATCTTTTGATGATA 421
```

```

1316 1316 ATCTCATGATCTTTGGCAATCTCGAGGAATTTCCCCCTATCACTCCAA 1365
461 His1ysAspPheLeuGluLeuLeuGlnAsnIleGluGlnGluGlnAspHis 477
1366 GATTCAACACTTTTCAAACTCTTCTCATAAAGATCACTCAAGAACAATAAAA 1415
477 sGln1ysArgSerGluLeuValSerGlnAlaSerLeuTyrLeuGluThrP 494
1416 ACACCTACGAAATACACTATATATGAGCCCTTGACTACTTTAGAGCACT 1465
494 heHis1LeuLeuGluProTleTyrHisAspAlaPheGlnPheAlaMetAsn 510
1466 GTCACATCTTCGAACCACTATGTCAATCAAAATCTTCGAATTCGTTGAAC 1515
511 Lys1ysLeuSerAsnLeuGlyValSerProThrGlyValValAspPheAr 527
1516 AAAACAACTATAAAACTTTAATCTTTTGTTCGACGAACTTCAGACTTCG 1565
527 gTyrAlaLysGlu 531
1566 TTTTATAGAAAAA 1578

.name: gb_ba:AE002215

documentation_block:
  1316 bp DNA linear
  Chlamydomophila pneumoniae AR39, section 43 of 94 of the
  genome.
  accession AE002215
  version AE002215.2
  GI:8163458

Chlamydomophila pneumoniae AR39.
Chlamydomophila pneumoniae AR39
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila
1 (bases 1 to 12980)
Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Uttel,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Genome sequences of Chlamydia trachomatis MoPn and C.M.
pneumoniae AR39
Nucleic Acids Res. 28 (6), 1397-1406 (2000)
20150255
10684935
2 (bases 1 to 12980)
Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Uttel,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Direct Submission
Submitted (01-MAR-2000) The Institute for Genomic Res
Medical Center Dr, Rockville, MD 20850, USA
On Jun 1, 2000 this sequence version replaced gi:7199
Location/Qualifiers
1. .12980
/organism="Chlamydomophila pneumoniae AR39"
/strain="AR39"
/db_xref="taxon:115711"
/note="synonym: Chlamydia pneumoniae AR39"
complement(197.1609)
/gene="CP0560".
complement(167.1609)
/gene="cp0560"
/note="sodium:sulfate symporter family proteo
by match to PRAM protein family HMM PF00939"
/codon_start=1
/trans_table=11
/product="sodium:sulfate symporter family proteo
protein_id="AAF73687.1"
/db_xref="GI:8163459"

```







179 ualaLeuProValPhePheProValHisLysSerGlnArgThrLeuGlns 191  
540 CGCTCGCCCGTATTCTCCCGCTTCTCATCACACCGCTAGGGAATCCTATA 589  
196 erLysSerLeuPro.....IleAlaSerGlyAlaPheTyPro 208  
590 AGAAGGAACACCCCATCCACATACATCTCCATGGCGCCTTGTCTTA 639  
209 LysAsnIleLysGlnLysGlnTrpLysLeuSerLysAsnProHisTy 225  
640 ARAAACATGAACACCAAACTACTTAATTTAGAAAAAATCCTCACTA 689  
225 rTyAsnGlnSerGlnValGluThrLysThrIleHisPheIleP 242  
690 CTATCATCAATCACTAAGTTAGCGAGTCACCTTAAAAATTATCC 739  
242 roAspAlaAsnThrAlaAlaLysLeuPheAsnGlnGlyLysLeuAsnTrp 258  
740 CAGACGCTCCACAGCACGAACCTTTCAAAAGATAAATCTATAGATTGG 789  
259 GlnGlyProTrpGlyGluArgIleProGlnGlnThrLeuSerAsnLe 275  
790 ATTGCTCACCTGGAGCGCTCGATATCTACAGAGAC...CAAAAAGT 836  
275 uGlnSerLysGlyHisLeuHisSerPheAspValaGlyThrSerTrpL 292  
837 TCTCTCCCAAGAAAGATTTCTACCTATTCTGTTCAGCACCACTTC 886  
292 euThrPheAsnIleAsnLysPheProLeuAsnAsnMetLysLeuArgGlu 308  
887 TTATCTATACCTGCAAAACCTCTAATACAAATAAAGCCCTCAGAAA 936  
309 AlaLeuAlaSerAlaLeuAspLysGluAlaLeuValSerThrIlePheLe 325  
937 GCATTCGCTCATGCTATTCATAGAAAATCTATCTTAAGACTCGTCCCTC 986  
325 uGlyArgAlaLysThrAlaAspHisLeuLeuProThrAsnIleHisSer 342  
987 AGGA.....CAAGAACTGTAACTCTAGTTCCCAAAATCTTTCACAAC 1030  
342 yrProGluHisGlnLysGlnGluMetAlaGlnArgGlnAlaTrpAlaLys 358  
1031 TCATCTTTCAAAGAGATCTCAACAGAAAGACGACAAACAAGCCAGA 1080  
359 LysLeuPheLysGluAlaLeuGluLysLeuGlnIleThrAlaLysAspLe 375  
1081 GCATATTTTCAAGACTAAGAACACTT.....TCTGAAAAAGAACT 1124  
375 uGluHisLeuAsnLeuIlePheProValSerSerAlaSerSerLeuL 392  
1125 CGCAGAACTCAGATCCTCTATCTCTATAGATTCTCGAATTCCTCCATCA 1174  
392 euValGlnLeuIleArgGluGlnTrpLysGluSerLeuGlyPheAlaIle 408  
1175 TAGTCTGAAGAAATCCAAGACAACTTAAAGATACCTTAGGATGGAAATC 1224  
409 ProIleValGlyLysGluPheAlaLeuGlnAlaAspLeuSerSerGI 425  
1225 AAAATCCAAGCATGGAGTACCCTGCTTTTTAAAGAAACGTCGTCGAAG 1274  
425 yAsnPheSerLeuAlaThrGlyGlyTrpPheAlaAspPheAlaAspProm 442  
1275 AGATTCTTCATAGCAGCAGGAGTGGATTGGGAATACGTGAAGCCCGC 1324  
442 etAlaPheLeuThrIlePheAlaTrpProSerGlyValProProTyAla 458  
1325 TAGCTTCTCTATCTATTCTAGGCAACCCAGAGACTCTCACAAATGG... 1371  
459 IleAsnHisLysAspPheLeuGluIleLeuGlnAsnIle.....GI 472  
1372 ...AGAAACAGTATTAGAAAGAACTTTAGAGAAGCTTATCTCCCTCA 1418

199 ACCCTTAGGAAA...CATCTCGTTATTGGCATGAGCCATGATCTGCCG 245

342 yrProGluHisGlnLysGlnGluMetAlaGlnArgGlnAlaTyrAlaLys 358  
 1131 TCATATCTTCAAAAGAGATCTCACAGAAAGACGACAAACAAAGCCAGA 1180  
 359 LysLeuPheLysGluAlaLeuGluGlnGlnIleThrAlaLysAspLe 375  
 1181 GCATATTTTCAAGAACTAAAGAACACTT.....TCTGAAAGAACT 1224  
 375 uGluHisLeuAsnLeuIlePheProValSerSerAlaSerLeuL 392  
 1225 CGCAGAACTCAGACTCTCTATCTATAGATTCTCGAATTCCTCATCA 1274  
 392 euValGlnLeuIleArgGluGlnIlePheLysGluSerLeuGluPheAlaIle 408  
 1275 TAGCTCAAGAACTCAAGAGAACTTAAAGATACCTTAGGATGAAATC 1324  
 409 ProIleValGlyLysGluPheAlaLeuGlnAlaAspLeuSerSerG 425  
 1325 AAAATCAAGGATGAGTACCCTGCTTTTAAAGAAACGTCGTCAAG 1374  
 425 YAsnPheSerLeuAlaThrGlyGlyTTPheAlaAspPheAlaAsp 442  
 1375 AGATTTCTCATAGCGACAGAGAGATGATGCGGAATACGTAAGCCCG 1424  
 442 etAlaPheLeuThrIlePheAlaTyrProSerGlyValProPheAla 458  
 1425 TAGCCTTCTCTATCTATTCTAGCAACCCAGACAGCTTACACATGG... 1471  
 459 IleAsnHisLysAspPheLeuGluIleLeuGlnAsnIle.....G 472  
 1472 ...AGAAACAGTGAATTCAGAAAGACTTTAGAGAACTCTATCTCCCTCA 1518  
 472 uGlnGluGlnAspHisGlnLysArgSerGluLeuValSerGlnAlaSerL 489  
 1519 TGCTTCAAGAGAAATTAACCGCAGAAATGATA..... 1555  
 489 euTyrLeuGluThrPheHisIleIleGluProIleTyrHisAspAla 505  
 1556 ....ATAGAAGAAACCCCGATTATCCCTCTGTATCAGCGCAATAT 1600  
 506 GlnPheAlaMetAsnLysLysLeuSerAsnLeuGlyValSerProThrG 522  
 1601 ATTACGTTATACATCTCTAAATCCAGATACATATCGGATCTCTTCAG 1650  
 522 yValValAspPheArg 527  
 1651 CCACACAGATCTCAA 1666  
 seq\_name: gb\_ba:AE002313

seq\_documentation\_block:  
 LOCUS AE002313 11402 bp DNA linear BCT 26-MAY-2000  
 DEFINITION Chlamydia muridarum, section 44 of 85 of the complete genome.  
 ACCESSION AE002313 AE002160  
 VERSION AE002313.1 GI:7190484

KEYWORDS  
 SOURCE Chlamydia muridarum.  
 ORGANISM Chlamydia muridarum  
 Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.  
 1 (bases 1 to 11402)  
 Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,  
 White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,  
 Berryman,C., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,  
 Bowman,C., Dodson,R., Gwin,M., Nelson,M., DeBoy,R., Kolonay,J.,  
 McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.  
 Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 pneumoniae AR39

TITLE  
 JOURNAL Nucleic Acids Res. 28 (6), 1397-1406 (2000)  
 MEDLINE 20150255  
 PUBMED 10684935  
 REFERENCE  
 2 (bases 1 to 11402)  
 Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F.,  
 White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,

Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,  
 Bowman,C., Dodson,R., Gwin,M., Nelson,M., DeBoy,R., Kolonay,J.,  
 McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.  
 Direct Submission  
 Submitted (01-MAR-2000) The Institute for Genomic Research, 9712  
 Medical Center Dr, Rockville, MD 20850, USA

# TITLE JOURNAL

## FEATURES source

1. 11402  
 /organism="Chlamydia muridarum"  
 /db\_xref="taxon:83360"  
 /note="synonym: Chlamydia trachomatis MoPn"  
 250..1569  
 /gene="TC0440"  
 250..1569  
 /gene="TC0440"  
 /note="similar to GP:2444077; identified by sequence  
 similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="phospholipase D family protein"  
 /protein\_id="AAF39294.1"  
 /db\_xref="GI:7190486"  
 /translation="MCSPCPRLHPIVTDCTCTPPLKRLPSLTAAAVVVQLFK  
 YHRKDLSHRYIYDYGKNTKPKTQSSNRHSLSATSSKTSRKLRLARRTGSSVP  
 RPIAQQLITFIQAQKTNNSPLKICKFIDRTHQLHIOVCFNHPTLIHLVKA  
 REVSVYVRDGTLEBEACKDSTICLOKSGRSLPHKSMGVDSORVLASNGTPOS  
 TEQDINLSLINDSELAHETNOVTPQIGNQFLYIPIFRNKDRLAVRLTKCI  
 AQAKTIVVVVILLHPVLASLQAAARGVKVEAVDTRSEQDTRLERLQSLPL  
 RVKPGAPRVHAKCYIDNKILICGSANSSAGLNRNREDFVIRGLTETQROSLSEI  
 WQSVKEKTEPLTAQSLKPREEDDPGEGTSSGISAGASAKAKTQ"  
 complement(1661..3049)  
 /gene="TC0441"  
 complement(1661..3049)  
 /gene="TC0441"  
 /note="hypothetical protein; identified by Glimmer2;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAF39295.1"  
 /db\_xref="GI:7190486"  
 /translation="MVDKFFFIQQLSLTSLSDDTCTDQFTKGLPKADIVHLP  
 GTISPEAWELGVNRGLIKWKNQWSPAFSGNPHKSVAEIRFNRESIRYERPNVS  
 LLKAITITDFASDFRIMATVQGHPPGIGIONDLWLQNLVQKIQODNLIIVTEV  
 QONRIAYVYPSLPLEARLYDLFASQLFLSOGITLFLNCFNKSGSNLAQS  
 TQORSEAAWLDERASFNNLFVQLSAGAESCPAPTKLTSGYHHATENGFGCEA  
 HAGEGTFLYLNQTLQALPIORAHGFOALEHVPTIHAIOEKNTLWMAFALNLVGA  
 SLHYSGLKISKTMINLDDHPFALFRDHLKSLTSDNPQMGTSVQNTMLLSG  
 FSADESHLTHVTSSPLTFKEIQLNVEATVSFVDVDTKIMLLNSIHEVLSYLVKIK  
 KQYNLLEGRKFL"  
 complement(3064..4602)  
 /gene="TC0442"  
 complement(3064..4602)  
 /gene="TC0442"  
 /note="similar to GB:M10101 SP:P04079 PID:146276 GB:U00096  
 PID:1788854; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="GMP synthase"  
 /protein\_id="AAF39296.1"  
 /db\_xref="GI:7190487"  
 /translation="MSNLIILDFGSOYTNVLAKKIRLLSVFCEVLPWNTPELKILOKS  
 PSLGLFSGPHSVVQNNSPEDVKIYNSNIPILGVCYGMQLIADRGSEVQGGKSEFG  
 YPIGVYSELSEKGLADQDAFHETRMSCDCVVPKDFVIASSOHCPIAAISCPK  
 KILFGLQHPVPSVQDGLKILNSFVHIQOTSETWKIETIEKLIQIRKVGETE  
 RVLLSGGVDSVSVLHNLALRSLCVFYDTGLLRKNEVEEVKQFSSILGTEIV  
 VDASEKFDHLSGDEPEQKRKVGIAFEVDEASKNLDVOWLAQCTIYSDVIESAK  
 SCDAQVIKSHNVGGPEKLNKLEPLRFLFDEVKALGKVLGLPDVILSRHPPG  
 PGLGVYLVGEVRYEYVEIKNADSIPIELKANLYHKVSOAFVLPCKSVKVGDC  
 RHYGTIALRAVESTDFMTACWPSUSREFLNCSRIINEIPEVCRVYDIDSKPPAT  
 LSW"  
 complement(4618..5691)  
 /gene="TC0443"

gene





7680 TCGGTTAGCATGTGCTCGGTTAAATAAAAAATTCATGCTGCTTCAAAAAG 7729

147 IyHisLeuSerIleAspHisPheGlyValHisSerProAsnGluSerThr 163  
|| :::  
7730 GAGACATTCCGTGGTGAANAATTAAGTGTTTCGC GCCTTTAGATGAACAACATACT 7779

164 LeuValValThrLeuGlnSerProThrSerHisPheLeuLysLeuLeuAl 180  
||| :  
7780 CTTGAAATCAGCTAGAACACCCTCTCTCTCATTTTTTAGAACAATCTCAC 7829

180 aleuProValPhePheProValHisLysSer.....GlnArgT 193  
||| :  
7830 ACATCCGTTTTTATCCAGTGCATGCTTCTTTGAGAGAATATTATGAA 7879

193 hrLeuGlnSerLysSerLeuProIleAlaSer...GlyAlaPheTyrPro 208  
|| :  
7880 ATCGAAGTAACGCTCTCTCCCAATAATTTCTAACGGCCCTTTATCATT 7929

209 LysAsnIleLysGlnLysGlnTrpIleLysLeuSerLysAsnProHisti 225  
::  
7930 CGGTGTTATGAGCCACAAATATTTTGTGTGCTGTATAAAATCCTTTTA 7979

225 rTyAsnGlnSerGlnValGluThrLysThrIleThrIleHisPheIleP 242  
|| :  
7980 TCATGACCAAAAAATGTTCTTTTAGATGCTGCTAGCTTCAAAATAGTTC 8029

242 roAspAlaAsnThrAlaAlaLysLeuPheAsnGlnGlyLysLeuAsnTrp 258  
|||| :  
8030 CCGATATTCATACAGCGGTGCAATGCTTTCAAAAAATATGCGGATTTA 8079

259 GlnGlyProThrPrpGlyGluArgIleProGlnGluThrLeuSerAsnLe 275  
||| ||| :  
8080 GTGGGGTGCCTTGGAGTCCTCTCTCCTCTAGAGAACAACAAAAATCT 8129

275 uGlnSerLysGlyHisLeuHisSerPheAspValAlaGlyThrSerTrpL 292  
|| :  
8130 T...TCTCAAGATTTTGTATGATATCCGGTGCTAAATGCACAGTTC 8176

292 euThrPheAsnIleAsnLysPheProLeuAsnMetLysLeuArgGlu 308  
||| :  
8177 TATCTCGCAATGTAACCAACAACCATTAGATATCCTTCGCTAGAGCG 8226

309 AlaLeuAlaSerAlaLeuAspLysGluAlaLeuValSerThrIlePheLe 325  
||| :  
8227 GCACCTCTTTAGCCATTGATCGAGAACAATATTAAAA....TTAGC 8270

325 uGlyArgAlaLysThrAlaAspHisLeuLeuProThrAsnIleHisSerT 342  
||| :  
8271 TGCAAAAGGAAGCATGTGTACCATGTTTGTACATCCTAGTCTATCTAGA 8320

342 yProGluHisGlnLysGlnGluMetalGlnArgGlnAlaTyrAlaLys 358  
||| :  
8321 TGCCCTTAGATGTTCTTTCGCAA...AAGAGAGAATCTCTTACTATAA 8367

359 LysLeuPheLysGluAlaLeuGluGluLeuGlnIleThrAlaLysAspLe 375  
|| :  
8368 AATACCTAGCGGAAGCTTTAAAA.....ACTGTCTCCTCAAGAAGAGTT 8411

375 uGluHisLeuAsnLeuIlePheProValSerSerSerAlaSerLeuL 392  
|| :  
8412 AAAAAAATCACACTCATCTACCCCATAGANTCTATTGTCTTGGCAGCAG 8461

392 euValGlnLeuIleArgGluGluTrpLysGluSerLeuGlyPheAlaIle 408  
||| :  
8462 TTGTTTCAAGATTCGCAACAGTTATTGATGTTTAGGATTTAGATT 8511

409 ProIleValGlyLysGluPheAlaLeuLeuGlnAlaAspLeuSerSergl 425  
|| :  
8512 TCTACTTTAGGATTTAGAAATACCACAGTTTTTTGGATAAAGCTTCTAAAG 8561

425 yAsnPheSerLeuAlaThrGlyGlyTyrPheAlaAspPheAlaaspprom 442  
|| :  
8562 AGAGTTTTCITTCACACAGGAANATGGGTGCTGATTTATCAACGCCA 8611

442	e tAlaPheLeuThrIlePheAlaTy rProSerGlyValProProTy rAla	458			
8612	AAGCTTTTCTATCCATCCTTGTAATGGACTAAATATAACATATCA A	8661			
459	Ile.....AsnHisLysASPheLeuGluIleLeuGlnAsn.....	470			
8662	GTGATTGATGGCAAATACAGAATATACGGATATCGTATCTGCAC TTT	8711			
471	.IleGluGlnGluInAspHisGlnLysArgSerGluLeuValSerGlnA	487			
8712	GTTGAGGATTCACAGATTTCACAAATCTACAGAA .....	8748			
487	laSerLeuTy rLeuGluThrPheHisIleIleGluProIleTy rHisAsp	503			
8749	..CAACTATTGCTCAAGAAATCCCTTTATT...CCCTTTATCACTG	8793			
504	AlaPheGlnPheAlaMetAsnLysLysLeuSerAsnLeuGlyValSerPr	520			
8794	GATTATCCTTAGCTTAACAATCTTAAGTCTCTAATCTCCAACCTCC TC	8843			
520	oThrGlyValValaspPheArg	527			
8844	ACTCGGAGAAATTGATTTAAA	8865			
seq_name: gb_ba:AE001291					
seq_documentation_block:					
LOCUS	AE001291	10827 bp	DNA	linear	BCT 30-OCT-2000
DEFINITION	Chlamydia trachomatis section 18 of 87 of the complete genome.				
ACCESSION	AE001291 AE001293				
VERSION	AE001291.1 GI:3228573				
KEYWORDS					
SOURCE	Chlamydia trachomatis.				
ORGANISM	Chlamydia trachomatis				
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.				
AUTHORS	Stephens,R.S., Kalman,S., Lammel,C.J., Fan,J., Marathe,R., Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q., Koonin,E.V. and Davis,R.W.				
TITLE	Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis				
JOURNAL MEDLINE	Science 282 (5389), 754-759 (1998)				
PUBMED	99000809				
REFERENCE	9784136				
AUTHORS	2 (bases 1 to 10827) Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W., Olinger, L., Grinstead, J., Davis, R. W. and Stephens, R. S.				
TITLE	Comparative genomes of Chlamydia pneumoniae and C. trachomatis				
JOURNAL MEDLINE	Nat. Genet. 21 (4), 385-389 (1999)				
PUBMED	99206606				
REFERENCE	10192388				
AUTHORS	3 (bases 1 to 10827) Stephens, R. S., Kalman, S., Lammel, C. J., Fan, J., Marathe, R., Aravind, L., Mitchell, W. P., Olinger, L., Tatusov, R. L., Zhao, Q., Koonin, E. V. and Davis, R. W.				
TITLE	Direct Submission				
JOURNAL	Submitted (20-MAY-1998) Program in Infectious Diseases, University of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA				
FEATURES	Location/Qualifiers 1..10827 /organism="Chlamydia trachomatis" /strain="D/UW-3/Cx" /db_xref="taxon:813" 206..1384 /gene="trpB" /note="Ctl70" 206..1384 /gene="trpB" /codon_start=1 /transl_table=11 /product="tryptophan Synthase (Beta Chain)" /protein_id="AAC67761.1"				
gene					
CDS					

```

/db_xref="GI:3328574"
/translation="MFKKHFGGAFLEPELLAQIOMKAEWELIKTOQSPFSELDCCI
LKNVAGQTEPLTEVKNFARAIDGPRVFLKREDLLHTGAHKLNALGCCILAKYLCKTR
VVAETGQHCQVATACAVLGLDCVYMGAKDVEROKPNYEMRGLGAEVVSIVTKGS
CGLADYNOALQDQWATHSTHYCLGSALGPLPYDIVERFFQSVISAEVKQIHAVAG
RDPDILLACIGGSNALGFRHFHPNPKVQLIGVEGGIGISSGKHAARFATPGPVF
HGFYSYLLQDDQVQLOTHSISAGLDYPSGDFHAENHESGRAFTLATDEALRAFF
LLTRNEGIIIFALESSHALAHVSIAPSLPEKQIIVNLSRGDKDLPQIIRNRGIVE
"
1377.2138
/gene="trpA"
/gene="CTI171"
1377.2138
/gene="trpA"
/codon_start=1
/transl_table=11
/product="Tryptophan Synthase (Alpha Chain)"
/protein_id="AAC67762.1"
/db_xref="GI:3328575"
/translation="MSKLIQFKQTKICIGYIAGDGGSTYIEAAKALIQGGVDILE
LGFPSDPVADNPEIQVSHDRALAEINTSETLIEIVGIRAFNQEVPLIYVYNPL
QRDLXYLRKLDAGINGVVIDLPAIPLSHGKSPFFEDILLAVGLDPLILLISAGTTPR
MSLQIYARGFLYIYPCQATRSEVGIKEBEKRVREHEDLPVDRRDICDKKEAAHVL
NYSDFIVTAFVHOT"
complement(2385..2876)
/gene="CTI172"
complement(2385..2876)
/gene="CTI172"
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67763.1"
/db_xref="GI:3328576"
/translation="MNYHNTFVKTSMPFLAKRIVOLNKNPFLKKFSETVLFIFERO
LKWMEYSIDENNYISDYNMEFGRPLIQIANPVCKALIQLEAQAWTILSNQWVG
DIYLVMSPISEKSVLLLETLINIIYQESLFLFKKPNVQCPKMSFEGHAILKIFL
TVS"
complement(2904..3074)
/gene="CTI172.1"
complement(2904..3074)
/gene="CTI172.1"
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67773.1"
/db_xref="GI:3522888"
/translation="MLRFLCLGYFPQLEKQKNLEDLIYNKLFVETVVMGAGCVV
NLAIGGGTILL"
complement(3142..3414)
/gene="CTI173"
complement(3142..3414)
/gene="CTI173"
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67764.1"
/db_xref="GI:3328577"
/translation="MKICRICDAIKRDPATFETERTYKNGSIVTSRQCPGVIRE
IGQKAILHNGEELQAVATGLKHLIIPISYKHSLLYKRLVT"
complement(3501..3956)
/gene="CTI174"
complement(3501..3956)
/gene="CTI174"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67765.1"
/db_xref="GI:3328578"
/translation="MLIFFDKSQSGALPDLERAGNLIRFVNRGMASQIKVTSQSG
HIFSEKMSVCKRIACIVLCIPLAPFCILGALIGITAYKLSYONALYLFREHRNM
CSEVEKAMGKNKQITRLQRNFRKVKKHADYKQKQYEQMCQSES"
4151..5740
/gene="oppA_2"

```

```

/note="CTI175"
4151..5740
/gene="oppA_2"
/codon_start=1
/transl_table=11
/product="Oligopeptide binding protein permease"
/protein_id="AAC67766.1"
/db_xref="GI:3328579"
/translation="MHRKFLAVSIAFVSIAFLGTLSCVHKKEPKDVLRIATICHDPMS
LDPROVFLSKDVSIVKALYECGLVREKAAAFOLAALAEVYHOSDDGVYTFELKNTFSN
GDWYAYDEESIKQIYFREIDNPISLSIALIKNSHAVLTCALGVEDLGVRLNKTLL
EIVLENPEPYELEILAHPPYVPHVSLREYTKDKRNRKRVPIISNGPFAICYEPQRY
LLINKNPLIHAKHDVLLNSVCLQIVPDHTAMQLFQKNHIDLVLGFWSSFSLEQORN
LPREKLFDPVLSVCLFNCNIHQITLNNPSLTALSLAINRETLKLAGKGCATSFFV
HPQLSQIPATTSIQDERIALAKGYLTKALTSQEDLEKITLIVPIESVCLAVQVEI
RQOLFDFVLGFKISTIGLEVHCFDKRSRGEFSLATGNWIADYHOASAFISLVLCNTRY
KDFOLINQONOKYINVAQLIQESSDQLMAEQLLLKESPLILPHLDYVYAKOPRV
SDLOFSSRGEDLKKRYSLAG"
complement(5767..6174)
/gene="dsbB"
/note="CTI176"
complement(5767..6174)
/gene="dsbB"
/codon_start=1
/transl_table=11
/product="Disulfide bond Oxidoreductase"
/protein_id="AAC67767.1"
/db_xref="GI:3328580"
/translation="MIKHLRSCLYLAWLFCIGTILMSVYYSVILNVEPCLCYORI
CLFLPVLIGIAAYREDISIKIYTIPLALNGFGIAIYQVCLQEIQWFLDICGVKSGS
TKLFLGFIPTMPWASAAAFACALLVLATKSK"
complement(6171..6887)
/gene="dsbG"
/note="CTI177"
complement(6171..6887)
/gene="dsbG"
/codon_start=1
/transl_table=11
/product="possible Disulfide Bond Chapterone"
/protein_id="AAC67768.2"
/db_xref="GI:6578099"
/translation="MDTRTPLRKKLIIISTALGFLVCVGLMHTKRSTMPKTHIPTT
AKVPTIGDPYAPINITYEPBSCSACEFSESEVPLIKKHEDVDTGEASLILVPCFI
RGMPPAAQLLCVYHHDKRPDPAYMEYFHRILTYKTKGSHWATPEVLAKAEKIP
THSREINLKGILQICNSRFEQLKANNIYQIMGQGLATPATAVGDVLIEDPTFD
EIKRVIITQRHQAIEEVR"
7033..8247
/gene="CTI178"
7033..8247
/gene="CTI178"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67769.1"
/db_xref="GI:3328582"
/translation="MENDFENVYHFRFRFKLSILLPALLGLMFFTPNMLYDYS
VILSDKICVILLILSALSYNPVVLQIGIFILAWFSFSSDLPPLFAHDSLIGF
ATLAILFLPNPEDELEVGTIPETCHYNGSGKRAAVLFAFVGLWQRYLTSAL
NIADADTSCSFFSSTLVIYVSMILVLSLTGGERKRWHTQPKVFTAILLFIAGLTL
AAILSQLPFLNRYEGVCLTVAPVFSLAFFVDETRATWYLSQYSNKQLATAPYGS
EYKESLFWERSVLPFSQACKQAPAGLSFPLNLIACITFSIFPVQJNANLSIPDTCR
FVNSACAFILVLSIFFAESLRLRWLSLLFAAGIILSPVIFHLPLEASTLISIVS
GIAFILLSIGRU"
complement(8249..8761)
/gene="CTI179"
complement(8249..8761)
/gene="CTI179"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67770.1"
/db_xref="GI:3328583"
/translation="MLAAGCAFLACIGGVLCICIDNVVFLSEIFLPLPILPILSLLLI

```



5689 CICTCTCTCGTGAGAAATTGATTAAAA 5716

seq\_name: gb\_ba:BACsp00K

seq\_documentation\_block:

LOCUS BACsp00K 5300 bp DNA linear BCT 06-MAR-1995  
 DEFINITION Bacillus subtilis spo0K operon.  
 ACCSSION M57689  
 VERSION M57689.1 GI:143602  
 KEYWORDS Spo0K operon; oligopeptide permease; sporulation protein.  
 SOURCE B. subtilis (strain JH642) DNA, clones pBR20/21, pJL2/3 and pJL7.  
 ORGANISM Bacillus subtilis  
 Bacteria; Firmicutes; Bacillus/Clostridium group;  
 Bacillus/Staphylococcus group; Bacillus.

REFERENCE 1 (bases 1 to 6300)  
 Rudner,D.Z., LeBeaux,J.R., Ireton,K. and Grossman,A.D.  
 TITLE The spo0K locus of Bacillus subtilis is homologous to the oligopeptide permease locus and is required for sporulation and competence

JOURNAL J. Bacteriol. 173 (4), 1388-1398 (1991)  
 MEDLINE 91139580

FEATURES  
 source

Location/Qualifiers  
 1..6300

/organism="Bacillus subtilis"

/strain="168"

/sub\_strain="JH642"

/sub\_species="Marburg"

/db\_xref="taxon:1423"

/clone="pJL7"

/clone="pBR20/21"

/clone="pJL2/3"

477..2114

/gene="spo0KA"

477..2114

/gene="spo0KA"

/codon\_start=1

/transl\_table=11

/product="sporulation protein"

/protein\_id="AAA62687.1"

/db\_xref="GI:143603"

/translation="MKKRWISVIMLIFTVLVLSAGFGSGSGEGKDKSGKTTNLI  
 NKTPEFSLHPLANDSVSGVIRQTPEGLTRINADGPEPGEMASKLETSGKNTTIF  
 TIRGQVSKNSGDPVTAODFEVANKWALDPNNSQAYQLYIKGAEAAANTGKSLDDV  
 AYKAVNDKTLKVELNPTPTFELTAFYIMPINKKIAEKKNKWTNAGDDVYVNSGPF  
 KWTAKSGSTLTKNDKDKVKKKIDKMMNNNTLKKFQAGELDWAGMPLG  
 QLPSTSLTKRGLSHVEPIAGVYKTEAKPLDNVIRKALTYSLDRQSIYKVN  
 TQGEQIPMAAAVPTMTKGFEDNKEGYFDKNDVKTAKVLEKGLKMSKASDLPIK  
 LSYTDDAHAKIAQAVQEMWKNLGVDELNSWNVYIDKLHSODYQIGRWGLDF  
 NDIINFLFRDKNGNDTGWENPEFKLLNQSTEDTKRAELLKRAEGIFIDEM  
 PVAPIYFVTDVWQDENLKGVIIMPGTGEVYFRNATFK"

477..536

/gene="spo0KA"

537..2111

/gene="spo0KA"

/product="sporulation protein"

222..3157

/gene="spo0KB"

222..3157

/gene="spo0KB"

/codon\_start=1

/transl\_table=11

/product="sporulation protein"

/protein\_id="AAA62688.1"

/db\_xref="GI:551726"

/translation="MLKYIGRLVYMIITLFIIVTVTFVFLMQAAGPGSPCEKKLPPE  
 EAILNNAHYGLDKPLFVQVYVYLSKVAWMDGPFYKYGQSVNDLISSGFPVSPFLGA  
 BAIIALALGVLFVIAALYHNKQWDTYVAILIIFGVSFIMAAYLVQVFSMKLGL  
 PFAVAGDSNAYTFLPSIALASMPMAFIARLSRSMIEVNSDIYTIKAKGLSAQRILQ  
 CGTFEFHCRLLHILGPNMAOVLGTSFIETIFIGPLGAHFVNSITNRDYTVIMVG  
 TVTFVSVILLICLVLDVLYGIIDPRIKLSKAKGA"

2255..3157

/gene="spo0KB"

/note="alternative translation start"

/codon\_start=1  
 /transl\_table=11  
 /product="sporulation protein"  
 /protein\_id="AAA62689.1"  
 /db\_xref="GI:143605"  
 /translation="MIITLFIIVTVTFVFLMQAAGPGSPCEKKLPPEIANLNAHYGL  
 DAPLFVQVYVYLSKVAWMDGPFYKYGQSVNDLISSGFPVSPFLGAEAAILALGV  
 LFGVIAALYHNKQWDTYVAILIIFGVSFIMAAYLVQVFSMKLGLFPAWQDSWAY  
 TFLPSIALASMPMAFIARLSRSMIEVNSDIYTIKAKGLSAQRILQCGTPEFHCR  
 LLHILGPNMAOVLGTSFIETIFIGPLGAHFVNSITNRDYTVIMVGTVTFVSVILLIC  
 VLIVDLYGIIDPRIKLSKAKGA"

gene

3161..4078  
 /gene="spo0KC"

3161..4078

/gene="spo0KC"

/codon\_start=1

/transl\_table=11

/product="sporulation protein"

/protein\_id="AAA62690.1"

/db\_xref="GI:143606"

/translation="MQNIPKMFEPAAANAGDAEKISKSLSLWKDAMLPPRSNKLAM  
 VGLIIVIIIMAFPMFESRYDYSTNLLNADKPPSKDHWGCTDGLGDIIVRTWVG  
 ARISIFIGVAAAVLDLLIGVINGSISGRGRTDEIMRTADILNAPSLMLVILLMV  
 VLPKGLFTIIITAMTITGWINMARIYRGVLOLKNOEYVLSOTLGAKTSRLFRHIVP  
 NAMGSLVITWLVPTAIFTEAFSLYSLGVPAFLASWGTMSDGLPALTYTPWRLEFF  
 PAGFICITMFGFNVGDLRDALDPLKLR"

gene

4083..5159  
 /gene="spo0KD"

4083..5159

/gene="spo0KD"

/codon\_start=1

/transl\_table=11

/product="sporulation protein"

/protein\_id="AAA62691.1"

/db\_xref="GI:143607"

/translation="MIRVTRLLEVKDLAISFKTYGGEVQAIRGVNPHLDKGETLAIIVG  
 ESSGSKSVTQALMKLIMPFGYKPKGEILFEGKDLVPLSERKEMQNVGKIGMIFOD  
 PNTSLNPTMKVQKITEVLFKHEKLSKEAKKRAVELLELVGIPMPERVNQPFHES  
 GMRQVYVAMALANPKLLIADBEFTALDVTIOAQILEMKDLQKIDTSIIFTHO  
 LGVVANVADYAVVAGQIVETVDEIFYDPRHPTWGLLASMPTLESSGEELTAI  
 PGTPDITNPKPGDAFALRSSYAMKIDFEQEPPEMKVSDTHYVKNLHPAPKVEPP  
 EAVKAMKRLANTFEKPVIVREGE"

gene

5152..6078  
 /gene="spo0KE"

5152..6078

/gene="spo0KE"

/codon\_start=1

/transl\_table=11

/product="sporulation protein"

/protein\_id="AAA62692.1"

/db\_xref="GI:143608"

/translation="MNETKLEIKHLKQHFVTPRGTVKAVDLSDFIVKGETLGIIV  
 GESGCGSTTGRSIIIRLYEATDGVLPNGENVHGRKSKKLLERKMKOMIFODPYAS  
 LNPRTVADIIAEGDLIHLAKTKKERMORVHELLETVGLNKEHANYRPHFSGGQRQ  
 RIGIARALAVDPEFIIDEPISALDSIQAVVNLKELQKEKGLTYLFIHADLSNVK  
 YISDRLGVMYFGKLVELAPADELNPHTKLSLSAIPLPDPDYERNVRQKYDPS  
 VHQLKDGTFMEFREVKPGHFVYMCTEAFKAFS"

CDS

5161..6078  
 /gene="spo0KE"

/note="alternative translation start"

/codon\_start=1

/transl\_table=11

/product="sporulation protein"

/protein\_id="AAA62693.1"

/db\_xref="GI:551727"

/translation="MTEKLEIKHLKQHFVTPRGTVKAVDLSDFIVKGETLGLVGES  
 GCGSTTGRSIIIRLYEATDGVLPNGENVHGRKSKKLLERKMKOMIFODPYASINP  
 RMTVADIIAEGDLIHLAKTKKERMORVHELLETVGLNKEHANYRPHFSGGQRQIG  
 IARALAVDPEFIIDEPISALDSIQAVVNLKELQKEKGLTYLFIHADLSNVKIS  
 DRIGMYFGKLVELAPADELNPHTKLSLSAIPLPDPDYERNVRQKYDPSVHQ  
 LKDGTFMEFREVKPGHFVYMCTEAFKAFS"

BASE COUNT

1919 a 1222 c 1462 g 1697 t  
 ORIGIN

seq\_name: qb\_pat:AX027646

```

seq_documentation_block:
LOCUS      AX027646                6300 bp      DNA
DEFINITION sequence 1 from Patent WO0039323.
ACCESSION  AX027646
VERSION     AX027646.1  GI:10188535
KEYWORDS   .
SOURCE      Bacillus subtilis.
            Bacillus subtilis
            Bacteria; Firmicutes; Bacillus/Clostridium group;
            Bacillus/Staphylococcus group; Bacillus.
REFERENCE  1 (bases 1 to 6300).
AUTHORS   Diaz-Torres, M. and Ferrari, E.
TITLE      Production of proteins in gram-positive microorganisms
JOURNAL    Patent: WO 039323-A 1 08-JUL-2000;
            DIAZ TORRES MARIA (US) ; FERRARI EUGENIO (US) ; GENENCOR INT (US)
FEATURES   Location/Qualifiers
            1..6300
            /organism="Bacillus subtilis"
            /db_xref="taxon:1423"
BASE COUNT 1919 a 1222 c 1462 g 1697 t
ORIGIN

```

```

alignment_scores:
    Quality: 617.50      Length: 540
    Ratio: 1.735        Gaps: 13
    Percent Similarity: 65.926   Percent Identity: 29.444

alignment_block:
US-09-824-567-2 x AX027646 ..

Align seg 1/1 to: AX027646 from: 1 to: 6300

10 lIeThrIlLeuLeuSerLeuSerValValLeuGlnGlyCysLysGlucose 26
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
498 GTCAAGTGTGATGCCTATTTCACCTCGTGCTGAGCGGTGGCGCTTGG 547
26 rSertHisSerThrSerArgGlyGlu..... 35
548 CGCGAGCGGATCAACGGTGAAGGMAAAGACAGTAAGAAGAACGACA 597
36 ..LeuAlaIleAsnIleArgAspGluProArgSerLeuAspProAlaGln 51
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
598 CACTTAACATAATATAAACTGAGCGGTTCTCTTCATCCGGCATGG 647
52 ValArgLeuLeuSerGluIleSerLeuValIleHisIleTyrGluGlyLe 68
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
648 GCAAATGATTCAGTATCAGCGCGGTGTTATCGTCGACAGACTTTTGAAGGATT 697
58 uValGlnGluAsnLeuSerGlyAsnIleGluProAlaLeuAlaGluA 85
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
698 GACACGT...ATCAATGCAGATGGTGAGCCGTGAGAAGGATGGCTTCTA 744
85 sPtyrSerLeuSerSeraspGlyLeuThrTyrThrPhelYsLeuLysSer 101
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
745 AAATTGAACGACGACGAGCAGGAAAGACATATACATTTACCATTCTGTAT 794
102 AlaPhe...TrpSerAsnGlyAspProLeuThrAlaGluAspPheIleGl 117
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
795 GGTTGGAATGCTTAATGGAGACCCCTGTAACCTGCCACAGATTTGAATA 844
117 uSerTrpLysGlnValAlaThrGlnGluValSerGlyIleTyrAlaPheA 134
TGCTTGGAAATGGGGCTTGACCCTAATAATGAATCACAAATACGCTTACC 894
134 laLeuAsnProTlleYsAsnValArgLysIleGlnGlyHisLeuSer 150
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
895 AGCTCTACTACATAAAGGTGCTGAAGCGGCAATACCGAAAAGGCAGC 944
151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValValTh 167
::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
945 CTAGACGATGTGGCAGTAAAGCTGTGATGTACAAACAGCTGAAGTTGA 994

```

[illegible]



```

1886 AGGCTGGGAAAAATCCAGAAATTCAGAAAGCTCTGAATCAGTCACAAACTG 1935
474 luGlnAspHisGlnLysArgSerGluLeuValSerGlnAla...SerLeu 489
1936 AAACAGATAAAACAAACACGTGCAGAGCTGCTCAAAAAAGCAGAAGGTATT 1985
490 TyrLeuGluThrPheHisIleIleGluProIleTyrHisAspAlaPheG 506
1986 TTCATTGATGAATGCCG...GTTGCCCAATCTATTTCATATCATGATAC 2032
506 nPheAlaMetAsnLysLysLeuSerAsnLeuGlyValSerProThrGlyV 523
2033 TTGGGTACAGAGATCAAAACCTAAAGGTGTTCATGCCAGGTACTIGTG 2082
523 alValAspPheArgTyrAla 529
2083 AGGTTTATTTCAGAAACGCA 2102
```

[illegible]

Align seg 1/1 to: AAD20238 from: 1 to: 1799

```

1 MetArgLysIleSerValGlyIleCysIleThrIleLeuLeuSerLeuSe 17
101 ATGCGCAAGATATCAGTGGGAATCTGTATCACCATTTCTCCTTAGCCTC 150
17 rValValLeuGlnGlyCysLysGluSerSerHisSerThrSerArg 34
151 CGTAGCTCTCCAGGCTCGAGGAGTCCAGTCACTCCTCTACATCTCGGG 200
34 lyGluLeuAlaIleAsnIleArgAspGluProArgSerLeuAspProArg 50
201 GAGAACTCCTATTATAATAGAGATCAACCCCGTCTTTAGATCCAGA 250
51 GlnValArgLeuLeuSerGluIleSerLeuValLysHisIleTyrGlu 67
251 CAGTGGGACTCTCTTCAGAAATCAGCCTGTCAAAACATATCTATGAGG 300
67 yLeuValGlnGlnAsnLeuSerGlyAsnIleGluProAlaLeuAlaG 84
301 ATTAGTCAAGAAATAATCTTCAGAAATATAGAGCTGCTCTTCGAG 350
84 luAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLysLeuLys 100
351 AAGACTACTCTCTTCTCGAGCGGACTCACTTATCTTTAAACTGAAA 400
101 SerAlaPheTrpSerAsnGlyAspProLeuThrAlaGluAspPheIle 117
401 TCAGCTTTTGGAGTAATGCGAGCCCTTAACAGCTGAAGCTTATAGA 450
117 uSerTrpLysGlnValAlaThrGlnGluValSerGlyIleTyrAlaPhe 134
451 ATCTTGGAAACAGTAGTACTACTCAAGAGTCTCAGGAATCTATGCTTTG 500
134 laLeuAsnProIleLysAsnValArgLysIleGlnGluGlyHisLeuSer 150
501 CTTTGAATCCANTAAATATGACAAAGATCCAGAGGGACACCTCTCC 550
151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValValTh 167
551 APAGACCATTTTGGAGTGCCTCTCTTAATGAATCTACACTTGTGTAC 600
167 rIleuGluSerProThrSerHisPheLeuLysLeuAlaLeuProValP 184
601 CTTGGATCCCCAACCTCCACTTTCTTAAACTTTTAGCTCTTCCAGTCT 650
184 hePheProValHisLysSerGlnArgThrLeuGlnSerLysSerLeuPro 200
651 TTTTCCCGGTTTCATAATCTCAAGAACCCCTGCATCCAAATCCAAATCTTACCT 700
201 IleAlaSerGlyAlaPheTyrProLysAsnIleLysGlnLysGlnTrp 217
701 ATAGCAAGCGAGCTTTCTATCTCTAAAAATATCAACAAAAACATGGAT 750
217 eLysLeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValGluThr 234
751 AAAACTCTCAAAAACCCCTCACTACTATAATCAAAAGTCAGTGGAAACTA 800
234 ystThrIleThrIleHisPheIleProAspAlaAsnThrAlaAlaLysLeu 250
801 AAAGATTACGATTCACTTCATCTCCGATGCAACACACAGCAGCAAAACTA 850
251 PheAsnGlnGlyLysLeuAsnTrpGlnGlyProProTrpGlyGluArg 267
851 TTTAATCAGGAAACTCAATTTGGCAAGGACCTCTTGGGGAGACGAT 900
267 eProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerP 284
901 TCCTCAAGAAACCTATCCAAATTTACAGTCTAAGGGGCACCTTACACTCT 950
284 heAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhePro 300

```

```

951 TTGATGTCCGAGGAACCTCATGCTCACTTCAATATCAATAAATTCCTCC 1000
301 LeuAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysG 317
1001 CTCACAAATATGAGCTTTAGAGAACCTTAGCATCAGCCTTAGATAAGA 1050
317 uAlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisL 334
1051 AGCTCTTGTCTCAACTATATTCTTAGGCCGTGCAAAACCTGCGCATCAT 1100
334 euLeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMet 350
1101 TCCTACCTACAAATATTCTAGTATCCCGAACATCAAAACACAGAGATG 1150
351 AlaGlnArgGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuGlu 367
1151 GCACAACGCCAGCTTACGCTTAAAAAACTCTTTAAAGAAGCTTTAGAAG 1200
367 uLeuGlnIleThrAlaLysAspLeuGluHisLeuAsnLeuIlePhePro 384
1201 ACTCAAATCACTGTCTAAGATCTCGAACATCTTAACTTATCTTTCCG 1250
384 aIserSerSerAlaSerSerLeuValGlnLeuIleArgGluGlnTrp 400
1251 TTTCTCTGTCAGCAAGTTCTTTACTAGTCCAACTTATACGAGACAGTGG 1300
401 LysGluSerLeuGlyPheAlaIleProIleValGlyLysGluPheAlaLe 417
1301 AAAGAAAGTTTAGGTTTCGCTATCCCTATTGTGCGAAAGAAATTTGCTCT 1350
417 uLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGly 434
1351 TCTCCAGCAGACCTTATCTTCAGGAACTTCTTTAGCTACAGAGAGAT 1400
434 rpPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePheAlaTyr 450
1401 GGTTCGACACTTTGCTGATCTATGCGATTTCTAACGATCTTTGCTTAT 1450
451 ProSerGlyValProTyrAlaIleAsnHisLysAspPheLeuGluI 467
1451 CCATCAGGAGTTCTCTTATGCAATCAACCATAGGACTTCTCTAGAGAA 1500
467 eLeuGlnAsnIleGluGlnGlnAspHisGlnLysArgSerGluLeuV 484
1501 TCTACAAACATAGAACAGCAAGATCCAAAAACGCTCGGAATTAG 1550
484 aIserGlnAlaSerLeuTyrLeuGluThrPheHisIleIleGluProIle 500
1551 TGTCCGACAGCTTCTCTTACCTAGAGACCTTTTCATATTATTGAGCCGATC 1600
501 TyrHisAspAlaPheGlnPheAlaMetAsnLysLysLeuSerAsnLeuG 517
1601 TACCACAGCAGCATTTCAATTTGCTATGAATAAAAACTTTCTTAATCTAG 1650
517 yValSerProThrGlyValValAspPheArgTyrAlaLysGluAsn 532
1651 AGTCTACCAACAGGAGTTGTGCACTTCGTTATGCTTAAGGAAAAAT 1696

```

seq\_name: /SDS5/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.AAX91990

seq\_documentation\_block:

ID AAX91990 standard; DNA; 1230025 BP.

XX  
AC AAX91990;

XX  
DT 13-SEP-1999 (first entry)

XX  
DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.

XX  
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope; ss.

OS Chlamydia pneumoniae.  
 PN WO9927105-A2.  
 PD 03-JUN-1999.  
 PF 20-NOV-1998; 98W0-IB01890.  
 PR 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.  
 PA (GBST ) GENSET.  
 XX Griffais R;  
 PI WPI; 1999-357842/30.  
 DR Genome sequence of Chlamydia pneumoniae  
 PT Claim 1; Page 291-611; 1912pp; English.  
 PS The present sequence represents the complete genome of Chlamydia  
 CC pneumoniae, and encodes proteins AY34584-Y35879. C. pneumoniae causes  
 CC respiratory disease such as pneumonia and bronchitis and is thought  
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,  
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides  
 CC encoded by the open reading frames of the C. pneumoniae genome (see  
 CC AY34584-Y35879) can be used in immunogenic compositions as vaccines.  
 CC Vectors containing C. pneumoniae nucleotide sequences can also be  
 CC used as immunogenic compositions, especially where the vector directs  
 CC the expression of a neutralising epitope of C. pneumoniae.  
 XX  
 S0 Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

alignment\_scores:  
 Quality: 2726.00 Length: 533  
 Ratio: 5.124 Gaps: 1  
 Percent Similarity: 99.812 Percent Identity: 99.812

alignment\_block:  
 US-09-824-567-2 x AAX91990 ..

Align seg 1/1 to: AAX91990 from: 1 to: 1230025

1 MetArgLysIleSerValGlyCysIleThrIleLeuLeuSerLeuSe 17  
 |||||  
 244480 ATGCGCAAGATATCAGTGGGAACTCTGTATCACCATTCTCTTAGCCCTC 244529  
 |||||  
 17 rValValLeuGlnGlyCysLysGluSerSerHisSerThrSerArg 34  
 |||||  
 244530 CGTAGTCCTCCCAAGGCTGCAAGGAGTCCAGTCACTCCTACATCTCGG 244579  
 |||||  
 34 lyGluLeuAlaIleAsnIleArgAspGluProArgSerLeuAspArg 50  
 |||||  
 244580 GAGAACTCGCTATTATTAAGATGAAGATGACCCCGTTCTTAGATCCAA 244629  
 |||||  
 51 GlnValArgLeuSerGluIleSerLeuValLysHisIleTyrGlu 67  
 |||||  
 244630 CAAGTCGACTCTTTTCAGAAATCAGCCTTGCAACATATCTATGAGG 244679  
 |||||  
 67 yLeuValGlnGluAsnLeuSerGlyAsnIleGluProAlaLeuAla 84  
 |||||  
 244680 ATTAGTTCAGAAATAATCTTTTCAGAAATATAGAGCTGCTCTTCAG 244729  
 |||||  
 84 luAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLysLeuLys 100  
 |||||  
 244730 AAGACTACTCTCTTCTTCGCGAGGACTCATTATATCTTTAAACTCAA 244779  
 |||||  
 101 SerAlaPheTrpSerAsnGlyAspProLeuThrAlaGluAspPheIle 117  
 |||||  
 244780 TCAGCTTTTGGAGTATGCGGCCCTTAAACAGCTGAAGACTTTATAGA 244829  
 |||||

117 uSerTrpLysGlnValAlaThrGlnGluValSerGlyIleTyrAlaPheA 134  
 |||||  
 244830 ATCTTGAACAAGTAGCTACTCAAGAAGTCTCAGAAATCTATGCTTTG 244879  
 |||||  
 134 laLeuAsnProIleLysAsnValArgLysIleGlnGluGlyHisLeuSer 150  
 |||||  
 244880 CCTTGAATCCAATTAATAATGCGAAAGATCCAAAGAGGACAGCTCTCC 244929  
 |||||  
 151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValValTh 167  
 |||||  
 244930 ATAGACCATTTTGGAGTCCACTCTCCTAATGAATCTACACTTGTGTAC 244979  
 |||||  
 167 rLeuGluSerProThrSerHisPheLeuLysLeuAlaLeuProValP 184  
 |||||  
 244980 CCTGGAATCCCAACCTCGCATTTCTTAAACATTTTAGCTCTCCAGTCT 245029  
 |||||  
 184 hePheProValHisLysSerGlnArgThrLeuGlnSerLysSerLeuPro 200  
 |||||  
 245030 TTTTCCCGTTCAATAATCTCAAGAACCCTGCAATCCAAATCTCTACCT 245079  
 |||||  
 201 IleAlaSerGlyAlaPheTyrProLysAsnIleLysGlnLysGlnTrpI 217  
 |||||  
 245080 ATAGCAACGGAGCTTCTATCTCTAATAATATCAAAACAACAATGGAT 245129  
 |||||  
 217 eLysLeuSerLysAsnProHisTyrTrpAsnGlnSerGlnValGluThrL 234  
 |||||  
 245130 AAACTCTCAAAAAACCTCACTACTATATAATCAAGTCAAGTGGAACTA 245179  
 |||||  
 234 yThrIleThrIleHisPheIleProAspAlaAsnThrAlaAlaLysLeu 250  
 |||||  
 245180 AARCATTTACGATTCATTCTCCGATGCAACACAGCAGCAAACTA 245229  
 |||||  
 251 PheAsnGlnLysLeuAsnTrpGlnGlyProProTrpGlyGluArgI 267  
 |||||  
 245230 TTTAATCAGGGAAAACTCAATTGGCAAGGACCTCTTGGGGAGAACCAT 245279  
 |||||  
 267 eProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerP 284  
 |||||  
 245280 TCCTCAAGAAACCCCTATCCAATTTTACAGTCTAAGGGGCACTTACACTCT 245329  
 |||||  
 284 heAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhePro 300  
 |||||  
 245330 TTGATGTGCGAGGAACCTCATGGCTCACCTTCAATATCAATAAATTCGCC 245379  
 |||||  
 301 LeuAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysG 317  
 |||||  
 245380 CTCACAATATGAAGCTTAGAAGCCCTTAGCATCAGCCTTAGATAAGA 245429  
 |||||  
 317 uAlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisL 334  
 |||||  
 245430 AGCTCTTGTCTCAACTATATTCTTAGGCCGTGCAAAACCTGCCGATCATC 245479  
 |||||  
 334 euLeuProThrAsnIleHisSerTyrProGluHisGlnLysGlnGluMet 350  
 |||||  
 245480 TCCTTACCTCAATAATATTCATAGCTATCCGACATCAAAAACAGAGATG 245529  
 |||||  
 351 AlaGlnArgGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuGlu 367  
 |||||  
 245530 GCACACGCCAGCTTACGCTAAAAAACTCTTTAAAGAGCTTTAGAAGA 245579  
 |||||  
 367 uLeuGlnIleThrAlaLysAspLeuGluHisLeuAsnLeuIlePheProV 384  
 |||||  
 245580 ACTCCAATTCACCGCTAAGATCTCGAACATCTTAATCTTATCTTTCGCC 245629  
 |||||  
 384 alSerSerSerAlaSerSerLeuLeuValGlnLeuIleArgGluGlnTrp 400  
 |||||  
 245630 TTTCTCTGTCAGCAAGTCTTCTTACTAGTCCAACTTATACGAGACAGTGG 245679  
 |||||  
 401 LysGluSerLeuGlyPheAlaIleProIleValGlyLysGluPheAlaLe 417  
 |||||  
 245680 AAAGAAAGTTTGGGTTGCTATCCCTATTGTCGGAAGGAATTTGCTCT 245729  
 |||||  
 417 uLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGlyT 434  
 |||||



```

228 lnSerGlnValGluThrLysThrIleHisPheIleProAspAla 244
    ::::||||| :::: :||| |||||:|||||
251445 AAGACGAGTACTGTACAGAGATCTGTATACATCATCTCGATCAA 251396
    ::::||||| :::: :||| |||||:|||||
245 AsnThrAlaAlaLysLeuPheAsnGlnGlyLysLeuAsnTrpGlnGlyPr 261
    :|||||: :||| :||| :||| :||| :||| :||| :||| :||| :|||
251395 CAAACTGCTTCTGCTTTATTCAACAGAGGAGCTAGATTGGCAAGTCT 251346
    :|||||: :||| :||| :||| :||| :||| :||| :||| :||| :|||
261 oProTrpGlyGluArgIleProGlnGluThrLeuSerAsnLeuGlnSerL 278
    :|||||: :||| :||| :||| :||| :||| :||| :||| :||| :|||
251345 CCCTGGGACATAGTATTCCGCAAGAACTTTAGCCACCAACAAAC 251296
    :|||||: :||| :||| :||| :||| :||| :||| :||| :||| :|||
278 ysGlyHisLeuHisSerPheAspValAlaGlyThrSerTrpLeuThrPhe 294
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
251295 GACGAGCTCCCAATCATTTGATATATCTGTGCTACTTCTGCTTACATTT 251246
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
295 AsnIleAsnLysPheProLeuAsnAsnMetLysLeuArgGluAlaLeuAl 311
    ||| ||| : : : : : : : : : : : : : : : : : : : : : : :
251245 AACCTGGCAAAAGCCCTTTTAGTCAATTCCTCAAGCTTCGCCAAGCTTGAG 251196
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
311 aserAlaLeuAspLysGluAlaLeuValSerThrIlePheLeuGlyArg. 327
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
251195 TCTAGTTTAAACAAAGACCTCTTCCCTTCCTGGCTTTGTNNAAACCT 251146
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
328 AlaLysThrAlaAspHisLeuLeuProThrAsnIleHisSerTyProG1 344
    ||||| ||||| : : : : : : : : : : : : : : : : : : : : :
251145 GCAGAA.....CATCTCTCTCTGTCACATTTGCACACCTACCCAGA 251105
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
344 whisGlnLysGlnGluMetAlaGlnArgGlnAlaTyAlaLysLysLeuP 361
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
251104 GCAGCTTCTTATAGCAAGAGCCATCCTTACCTTAACTTTAC 251055
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
361 heLysGluAlaLeuGluGluLeuGlnIleThrAlaLysAspLeuGluHis 377
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
251054 TAGAAGAAGCTCTGACTGAGCTTACATGACTATTAGGATCTAGAGAA 251005
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
378 LeuAsnLeuIlePheProValSerSerAlaSerSerLeuLeuValG1 394
    ||| ||| : : : : : : : : : : : : : : : : : : : : :
251004 TATCTCTTACCTTTTCCGACAGCTCTACTGTGAACCTCACAGATAGTCA 250955
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
394 nLeuIleArgGluGlnTrpLysGluSerLeuGlyPheAlaIleProIleV 411
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
250954 GATGTGCGGATCAGTGGCGAGAGTTTAGGAATTACTTCCCTTACT 250905
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
411 alGlyLysGluPheAlaLeuLeuGlnAlaAspLeuSerSerGlyAsnPh 427
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
250904 GCGGAAAGAATAATGCTTTGTTGAAACGATCTAATAGCAATACTTTC 250855
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
428 SerLeuAlaThrGlyGlyTrpPheAlaAspPheAlaAspProMetAlaPh 444
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
250854 TTTATGCTATAGTGGTGGTGGTTCGACGCTTCTGACCCCTTTAGCGTT 250805
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
444 eLeuThrIlePheAlaTyProSerGlyValProProTyAlaIleAsnH 461
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
250804 TCTTCCATTTTCTCC...TCGAAAGAGTCAACACTTATGCTTTACAG 250758
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
461 isLysAspPheLeuGluLeuLeuGlnAsnIleGluGlnGlnAspHis 477
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
250757 ATCTCAATTTGATCACTGCTCTCTATAGAACGGAACCAACCTT 250708
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
478 GlnLysArgSerGluLeuValSerGlnAlaSerLeuTyLeuGluThrPh 494
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
250707 CAAAAGCTCAGCTTTATTCGAGAGCTTCTCTATACATAGAGAGACA 250658
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
494 eHisIleLeuGluProIleTyHisAspAlaPheGlnPheAlaMetAsnL 511
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
250657 AAGCGCATAGAACCCCTCTACAGAGCTGTTCCTCATATACAAATA 250608
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
511 ysLysLeuSerAsnLeuGlyValSerProThrGlyValValAspPheArg 527
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
250607 ATAACTTTCTTTTGTAGACTACATCTTCAGGCGCTAGTTGATATCGG 250558
    : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

528 TyrAlaLysGluAsn 532
    |||||: :||| :||| :||| :||| :||| :||| :||| :||| :|||
250557 TATGCTAAAAACTCT 250543

```

```

seq_name: /SDS5/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:AA518750

```

```

seq_documentation_block:
ID AAS18750 standard; DNA; 1787 BP.

```

```

XX AAS18750;

```

```

XX DT 26-MAR-2002 (first entry)

```

```

XX Chlamydothila pneumoniae gene encoding an ATP-binding cassette.
XX ATP binding cassette; secretory locus open reading frame; endopeptidase;
XX secretory locus ORF; protease; metalloprotease; Clp protease APrase;
XX Clp protease subunit; transglycolase/transpeptidase; ClpC protease;
XX thioedoxin; Chlamydia infection; antibacterial; ds.

```

```

XX Chlamydothila pneumoniae CWL029.

```

```

XX WO200185972-A2.

```

```

XX 15-NOV-2001.

```

```

XX 08-MAY-2001; 2001WO-CA00653.

```

```

XX 08-MAY-2000; 2000US-202672P.

```

```

XX 30-MAY-2000; 2000US-207852P.

```

```

XX 16-JUN-2000; 2000US-211796P.

```

```

XX 16-JUN-2000; 2000US-211797P.

```

```

XX 16-JUN-2000; 2000US-211798P.

```

```

XX 16-JUN-2000; 2000US-211801P.

```

```

XX 16-JUN-2000; 2000US-212044P.

```

```

XX 26-SEP-2000; 2000US-235335P.

```

```

XX 26-SEP-2000; 2000US-235361P.

```

```

XX 26-SEP-2000; 2000US-235398P.

```

```

XX (AVET) AVENTIS PASTEUR LTD.

```

```

XX Murdin AD, Comen RP, Wang J, Dunn P;

```

```

XX WPI; 2002-049447/06.

```

```

XX P-PSDB; RAU09430.

```

```

XX Vaccine useful for immunising mammals against chlamydia infections,
XX comprises vectors having sequences of ATP binding cassette gene,
XX secretory locus open reading frame gene of chlamydia

```

```

XX Claim 1; Fig 1; 35pp; English.

```

```

XX The present invention relates to the isolation of Chlamydothila
XX pneumoniae strain CWL029 genes and their encoded proteins. The genes of
XX the invention encode an ATP binding cassette gene, a secretory locus
XX open reading frame (ORF), an endopeptidase, a protease, a
XX metalloprotease, Clp protease APrase, a Clp protease subunit, a
XX transglycolase/transpeptidase, a ClpC protease, or thioedoxin. The
XX genes of the invention can be used in a vector as a vaccine for the
XX prevention and treatment of Chlamydia infections. Also described are
XX B- and T-cell epitopes from the proteins of the invention which can be
XX used as Chlamydia antigens. AAS18750-AAS18759 represent the C. pneumoniae
XX genes of the invention.

```

```

XX Sequence 1787 BP; 628 A; 436 C; 261 G; 462 T; 0 other;

```

```

XX alignment_scores:

```

```

XX Quality: 755.50 Length: 522

```

```

XX Ratio: 2.116 Gaps: 9

```

```

XX Percent Similarity: 68.391 Percent Identity: 33.908

```

```

XX alignment_block:

```



US-09-824-567-2 x AAS18750

Align seg 1/1 to: AAS18750 from: 1 to: 1787

13 LeuLeuSerLeuSerValValLeuGlnGlnGlycylSerSerHis 29  
 149 CTTCCTTTCTTATGCTCCTCTTCAAGCTGCTCAAGCAAAACAGA 198  
 29 rSerThrSerArgGlyGluLeuAlaIleAsnIleArgAspGluProArg 46  
 199 ACCCTTAGAGAAA...CAATCGCTTATGCGATGAGCCATGATCGCG 245  
 46 erLeuAspProArgGlnValArgLeuLeuSerGluIleSerLeuValys 62  
 246 ACCTAGATCCTCGCAATCCCTATTAAAGCAGAGATGCTTCCCTAGCAAA 295  
 63 HisIleTyrGluGlyLeuValGlnGlnAsnLeuSerGlyAsnIleG 79  
 296 GCCTCTATGAAGAGACTGACAAGAGAACTGAT.....CAAGGATCGC 339  
 79 uProAlaLeuAlaGluAspTyrSerLeuSerSerAspGlyLeuThrTyr 96  
 340 ACTGGCTCTTCGAGAAGTATACCTGTCAAGATCATAGTCTATA 389  
 96 hrPhelLysLeuLysSerAlaPheTrpSerAsnGlyAspProLeuThrAla 112  
 390 CCTTAACTCAGACCTCTCTGTGGAGCGATGGCACTCCACTCCTGCT 439  
 113 GluAspPheIleGluSerTrpLysGlnValAlaThrGlnGluValSerG 129  
 440 TATGACTTTGAAAATCTATAAACAACACTGTACTTCGAAGAATTTTCAAC 489  
 129 yIleTyrAlaPheAlaLeuAsnProIleLysAsnValArgLysIleGln 146  
 490 TTCCATACATACTTACTTCGGCGTGATTAATAATTTCTTCGGCAATCCACA 539  
 146 luGlyHisLeuSerIleAspHisPheGlyValHisSerProAsnGluSer 162  
 540 ATGCTCAAAATCTCTGGAACCTCTGGGATACAGGCAAAAGATGATCTT 589  
 163 ThrLeuValValThrLeuGluSerProThrSerHisPheLeuLysLeu 179  
 590 ACTTTGGTGATACCTCAGCAACCTTTCCCATACTTCTCACACTTAT 639  
 179 uAlaLeuProValPhePheProValHisLysSerGlnArgThrLeuGln 196  
 640 CGCTCGCCCGTATCTCCCTGTTCATCACCCCTTAGGGAATCCTATA 689  
 196 erLysSerLeuPro.....IleAlaSerGlyAlaPheTyrPro 208  
 690 AGAAGAGAACACCCCATCATATCTCCAAATGGGCCCTTTGCTTA 739  
 209 LysAsnIleLysGlnLysGlnTrpIleLysLeuSerLysAsnProHis 225  
 740 AAAAACAATGACACCAAACTACTTAATTTAGAAAAAATCCTCATA 789  
 225 rTyrAsnGlnSerGlnValGluThrLysThrIleHisPheIleP 242  
 790 CTATGATCATGATCATGTAAGTATAGCCGAGTACCTCACTTAAATTTATCC 839  
 242 roAspAlaAsnThrAlaAlaLysLeuPheAsnGlnGlyLysLeuAsnTrp 258  
 840 CAGACGCTCCACAGCCAGCAAACTTTTCAAAAGTAATCTATAGATTGG 889  
 259 GlnGlyProProTrpGlyGluArgIleProGlnThrLeuSerAsnLe 275  
 890 ATTGGCTCCTTGGAGCGCTCCGATATCTAACGAAGAC...CAAAAAGT 936  
 275 uGlnSerLysGlyHisLeuHisSerPheAspValAlaGlyThrSerTrpL 292  
 937 TCTCTCCAGAAAAGATCTTACCTATCTGTTCGATCAAGCACCCTTC 986  
 292 euThrPheAsnIleAsnLysPheProLeuAsnAsnMetLysLeuArgGlu 308

987 TTATCTATAACCTGCAAAACCTCTAATACAAATTAAGCCCTCAGGAAA 1036  
 309 AlaLeuAlaSerAlaLeuAspLysGluAlaLeuValSerThrIlePheLe 325  
 1037 GCCATTGCTCATGCTATTGATAGAAAATCTATCTTAAGACTCGTGCCITC 1086  
 325 uGlyArgAlaLysThrAlaAspHisLeuLeuProThrAsnIleHisSer 342  
 1087 AGGA.....CAAGAAAGCTGTAACCTAGTTCCCCCAATCTTTCAACA 1130  
 342 yTrpGluHisGlnLysGlnGluMetAlaGlnArgGlnAlaTyrAlaLys 358  
 1131 TCAATCTTTCAAAAGAGATCTCAACAGAGAGACGACAAACAAGCCAGA 1180  
 359 LysLeuPheLysGluAlaLeuGluLeuGlnIleThrAlaLysAspLe 375  
 1181 GCATATTTTCAAGAGCTAAAGAACACTT.....TCTGAAAAAGAACT 1224  
 375 uGluHisLeuAsnLeuIlePheProValSerSerSerAlaSerSerLeu 392  
 1225 CGCAGAACTCAGATCTCTATCTCTATAGATTTCTCGAATTCCTCCATCA 1274  
 392 euValGlnLeuIleArgGluGlnTrpLysGluSerLeuGlyPheAlaIle 408  
 1275 TAGCTCAAGAAATCCAAAGACAACTTAAGATACCTTAGATTGAAATC 1324  
 409 ProIleValGlyLysGluPheAlaLeuLeuGlnAlaAspLeuSerG 425  
 1325 AAATCCAAAGGATGGATGACCTGCTCTTTTAAAGAAACGTCGCAAGG 1374  
 425 yAsnPheSerLeuAlaThrGlyGlyTrpPheAlaAspPheAlaAspPro 442  
 1375 AGATTCTTCATAGCAGAGGATGGATTGGGAATACGTAAGCCCG 1424  
 442 etAlaPheLeuThrIlePheAlaTyrProSerGlyValProProTyrAla 458  
 1425 TAGCTTCTCTATCTATTCTAGGCAACCCAGAGACCTCACACAATGG... 1471  
 459 IleAsnHisLysAspPheLeuGluIleLeuGlnAsnIle.....G 472  
 1472 ...AGAAACAGTGATTACGAAAAGACTTTAGAGAAACTCTATCTCCCTCA 1518  
 472 uGlnGluGlnAspHisGlnLysArgSerGluLeuValSerGlnAlaSerL 489  
 1519 TGCTTCAAAAGAGAATTTAAACGCGCAGAAATGATA..... 1555  
 489 euTyrLeuGluThrPheHisIleIleGluProIleTyrHisAspAlaPhe 505  
 1556 ....ATAGAAAGAAACCCGATTTATCCCTGTATCAGGCAAAATAT 1600  
 506 GlnPheAlaMetAsnLysLysLeuSerAsnLeuGlyValSerProThrG 522  
 1601 ATTACGCTATACCTCTAAATCCAGAAATACATTCGGATCTCTCTTAGG 1650  
 522 yValValAspPheArg 527  
 1651 CCACACAGATCTCAAA 1666

seq\_name: /SID5/gcdata/geneseq/geneseq-emb1/NA2000.DAT:AAA51864

seq\_documentation\_block:

ID AAA51864 standard; DNA; 6300 BP.

XX AAA51864;

AC

XX

DT 31-OCT-2000 (first entry)

XX B. subtilis opp operon.

DE

XX

XX

KW Opp operon; Spo0K; oligopeptide permease; sporulation; ABC transporter;

KW ATP-binding cassette transporter; mutation; protein production; oppA;

KW oppB; oppC; oppD; oppF; ds.

XX OS Bacillus subtilis.

XX FH Key Location/Qualifiers

FT CDS 477..2110

FT /tag= a

FT /label= oppA

FT /product= ligand-binding\_protein

FT /trans\_except= (pos:2106..2107,aa:Phe)

FT /note= "this codon contains an apparent 1 base deletion

FT which alters the reading frame"

FT CDS 222..3157

FT /tag= b

FT /label= oppB

FT /product= membrane\_protein

FT CDS 3161..4078

FT /tag= c

FT /label= oppC

FT /product= membrane\_protein

FT CDS 4083..5159

FT /tag= d

FT /label= oppD

FT /product= ATPase

FT CDS 5152..6078

FT /tag= e

FT /label= oppF

FT /product= ATPase

XX WO200039323-A2.

XX 06-JUL-2000.

XX 21-DEC-1999; 99WO-US31010.

XX 24-DEC-1998; 98GB-0028711.

XX (GENV) GENENCOR INT INC.

XX (DIAZ/) DIAZ-TORRES M.

XX (FERR/) FERRARI E.

XX Diaz-Torres M, Ferrari E;

XX WPI; 2000-452412/39.

XX P-PSDB; AAY97040, AAY97041, AAY97042, AAY97043, AAY97044.

XX Production of proteins, such as hormones, enzymes, growth factors or

XX cytokines, in gram-positive microorganisms containing a mutation in at

XX least one of the genes of the opp operon gene cluster

XX Disclosure; Fig 1A-M; 32pp; English.

XX The opp operon of Bacillus (also known as spoOK operon) encodes an

XX oligopeptide permease that is required for the initiation of sporulation

XX and the development of genetic competence. The opp operon is a member of

XX the family of ATP-binding cassette (ABC) transporters involved in the

XX import or export of oligopeptides from 3-5 amino acids. Bacillus strains

XX containing a mutation in the opp operon produce more recombinant protein

XX than the wild-type strain. Therefore, gram-positive microorganisms,

XX especially Bacillus strains, containing a mutation in at least one of the

XX genes of the opp operon can be used for heterologous protein production,

XX especially hormones, enzymes (preferably proteases such as subtilisin),

XX growth factors or cytokines.

XX Sequence 6300 BP; 1919 A; 1222 C; 1462 G; 1697 T; 0 other;

alignment\_scores:

Quality: 617.50 Length: 540

Ratio: 1.735 Gaps: 13

Percent Similarity: 65.926 Percent Identity: 29.444

alignment\_block:

US-09-824-567-2 x AAA51864 ..

Align seg 1/1 to: AAA51864 from: 1 to: 6300

10 IleThrIleLeuLeuSerLeuSerValValLeuGlnGlyCysLysGluSe 26

498 GTACGTTGATGCTCAATTTTCACTCTCTGCTGAGCGGCTGGCTTGG 547

26 rSerHisSerThrSerArgGlyGlu..... 35

548 CGGCGCGGATCAACAGCGTGAAGGAAAGGACAGTAAAGGAAAGACGA 597

36 ..LeuAlaIleAsnIleArgAspGluProArgSerLeuAspProArgGln 51

598 CACATTAACATTAATATAAAGTGAAGCGTCTCTCTTACATCCGGGATTG 647

52 ValArgLeuLeuSerGluIleSerLeuValLysHisIleTyrGluGlyLe 68

648 GCAAAATGATTTCAGTATCAGCGCGGTATTCGCTGAGAGGATTTTGAAGGAT 697

68 uValGlnGluAsnAsnLeuSerGlyAsnIleGluProAlaLeuAlaGluA 85

698 GACAGCT...ATCAATCCAGATGGTGAAGCGTGAAGAGGATGGCTTCTA 744

85 spTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLysLeuLysSer 101

745 AAATTGAACGACGACGAGGAGGAGACATATACATTTACCATTCTGTGAT 794

102 AlaPhe...TrpSerAsnGlyAspProLeuThrAlaGluAspPheIle 117

795 GGTGTGAATGGTCTAATGGAGACCGCTGTAACGCACAGATTTTGAATA 844

117 uSerTrpLysGlnValAlaThrGlnGluValSerGlyIleTyrAlaPheA 134

845 TCCTTGGAAATGGCGCTTGAACCTAATAATGAATCACAATACGCTTACC 894

134 LaLeuAsnProIleLysAsnValArgLysIleGlnGluGlyHisLeuSer 150

895 AGCTCTACTACATAAAAGGTGCTGAAGCGGCAATACCGGAAAGGACG 944

151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValValTh 167

945 CTAGCAGATGTGCGCAGTAAAGCTGTCAATGCACAAACGCTGAAGGTTGA 994

167 rLeuGluSerProThrSerHisPheLeuLysLeuLeuAlaLeuProValP 184

995 ATAAATAACCGGACTCCATATTTCACTGAATTAAGTGGGTTCTATACGT 1044

184 hePheProValHisLysSerGlnArgThrLeuLysSerLysSerLeuPro 200

1045 ATATGCGGATCAATGAG...AAAATTCGAGAGAAAATAAAGTGAAT 1091

201 IleAlaSerGlyAlaPheTyrProLysAsn.....IleLysGlnLysG 215

1092 ACAAAATGCGGAGATGATTAATGATCAACGCGCGCTCAAAATGACGCG 1141

215 nTrp.....IleLysLeuSerLysAsnProHisTyrTyrA 227

1142 ATGAAACACACGCGCTCTATTACTCTCGAAAAAATGACCAGTATTGGG 1191

227 snGlnSerGlnValGluThrLysThrIleThrIleHisPheIleProAsp 243

1192 ATAGACAAAAGTCAAACTGAAGAAAATCGATATGTTATGATCAACAAT 1241

244 AlaAsnThrAlaAlaIleLysLeuPheAsnGlnGlyLysLeuAsnTrpGln 1260

1242 ACAATACGGAACATAAAATAATCCAAAGCTGCGGAACCTGATGGCGCG 1291

260 yProProTrpGlyGluArgIleProGlnGluThrLeuSerAsnLysGln 277

1292 TATGCGGCTCGGACAG...CTTCCGACAGAACTCCCTCGCGACCTGAAA 1338

277 erLysGlyHisLeuHisSerPheAspValAlaGlyThrSerTrpLeuThr 293

1339 AAGACGGTCTTTTACATGTGACCCGATGACGAGTGTATTGGTACAAA 1388  
 294 PheAsnIleAsnLysPheProLeuAsnMetLysLeuArgGluAlaLe 310  
 ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||  
 1389 TTCAACACTGAAGCTAAGCCATAGACACGCTCAATATCGTAAACGCTT 1438  
 310 uAlaSerAlaLeuAspLysGluAlaLeuValSerThrIlePheLeuGlyA 327  
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 1439 AACATATCGCTTGACGCTGACGCTGACGCTGACGCTGACGCTGACG 1488  
 327 t9AlaLysThraLysPheHisLeuLeuProThrAsnIleHisSerTyrPro 343  
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 1489 AGCAAAATCCCGCAATAGCGTGCAGTGGCGCTACATGAGGATTTGAG 1538  
 344 GluHisGlnLysGlnGluMetAlaGlnArgGlnAla...TyrAlaLysLY 359  
 :  
 1539 GATACCAAGAGGATGACTTCAAGACAAATGCTCAAAACACAGCAAGA 1588  
 359 sLeuPheLysGluAlaLeuGluGluLeuGlnIleThr...AlaLysAspL 375  
 :  
 1589 ATACCTTGAAAAGGCGCTAAGAAATGGCGCTTAAGCAAGCATCTGATT 1638  
 375 euGluHisLeuAsnLeuLeuPheProValSerSerSerAlaSerSerLeu 391  
 || : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 1639 TGCAAAATCAATGCTTCTTACAC...ACTGATGACGACACACGCGAAA 1685  
 392 LeuValGlnLeuIleArgGluGlnTyrLysGluSerLeuGlyPheAlaI 408  
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 1686 ATGCTCAGCAGTACAAAGAAATGTGAAGAAATTTAGCGTGTGATGT 1735  
 408 eProLysValGlyLysGluPheAlaLeuLeuGlnAlaAspLeuSerSerg 425  
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 1736 TGAGCTTGATACTCAGAGTGGATGCTATATGATGATGATGATGATGAT 1785  
 425 LysAsnPheSerLeuAlaThrGlyClyTyrPheAlaAspPheAlaAspPro 441  
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 1786 AGATTATCAATGCGCGTATGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1835  
 442 MetAlaPheLeuThrIlePheAlaTyrProSer...GlyValProTyr 457  
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 1836 ATCAACTTCCTTGATTTCCCGCAAAACAGGAGGAAATAACGATAC 1885  
 457 rAlaIleAsnHisLysAspPheLeuGluLeuGlnAsnIleGluGlnG 474  
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 1886 AGCTGGGAAATCCAGAAATTCAAAAGCTTCTGAATCAGTCACAACTG 1935  
 474 LuGlnAspHisGlnLysArgSerGluLeuValSerGlnAla...SerLeu 489  
 || : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 1936 AAACAGATAAACAAACAGTGGCAGCTGCTGAAAAAGCAGAGGTATT 1985  
 490 TyrLeuGluThrPheHisIleLeuGluProIleTyrHisAspAlaPheG 506  
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 1986 TTCATTGATGAATGCGG...GTGCCCATCTATTCTTATCTACTGATAC 2032  
 506 nPheAlaMetAsnLysLysLeuSerAsnLeuGlyValSerProThrGlyV 523  
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 2033 TTGGGTACAGATGAACAACTTAAAGGTGTTATCATCGCAGGACTGCTGG 2082  
 523 alValAspPheArgTyrAla 529  
 || : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 2083 AGTTTATTTCAGAACGCA 2102

seq\_name: /SIDS5/gcgdata/geneseq/emb1/NA1999.DAT.AAX61764

seq\_documentation\_block:

ID AAX61764 standard; DNA; 1536 BP.

AC AAX61764;

XX

DT 19-JUL-1999 (first entry)

DE B. burgdorferi antigenic protein coding sequence, t606.nt.

XX

KW Antigenic protein; vaccine; Lyme disease; infection; detection; ss.  
 XX Borrelia burgdorferi.  
 OS WO9859071-A1.  
 PN 30-DEC-1998.  
 PD 18-JUN-1998; 98WO-US12718.  
 XX 03-SEP-1997; 97US-0057483.  
 PR 20-JUN-1997; 97US-0050359.  
 PR 22-JUL-1997; 97US-0053344.  
 PR 22-JUL-1997; 97US-0053377.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA (MEDI-) MEDIMUNE INC.  
 XX Choi GH, Erwin AL, Hanson MS, Lathigra R;  
 XX WPI; 1999-189980/16.  
 DR P-PSDB; AAY20067.  
 XX New isolated Borrelia burgdorferi nucleic acids - used to develop  
 PT products for the diagnosis, prevention and treatment of diseases  
 PT caused by Borrelia, particularly Lyme disease  
 XX Claim 1; Page 182-183; 275pp; English.  
 XX This sequence encodes a Borrelia burgdorferi (Bb) protein of the  
 CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
 CC can be used in vaccines for eliciting protective antibodies to members of  
 CC the Borrelia genus, particularly for the use against Lyme disease in  
 CC humans and animals. They can be used for preventing or attenuating an  
 CC infection caused by a member of the Borrelia genus. The products can also  
 CC be used for detection of members of the Borrelia genus.  
 XX Sequence 1536 BP; 631 A; 246 C; 256 G; 403 T; 0 other;

alignment\_scores:

Quality: 505.50 Length: 485  
 Ratio: 1.620 Gaps: 15  
 Percent Similarity: 64.330 Percent Identity: 28.041

alignment\_block:

US-09-824-567-2 x AAX61764 ..

Align seg 1/1 to: AAX61764 from: 1 to: 1536

38 IleAsnIleArgAspGluProArgSerLeuAspProArgGlnValArgLe 54  
 ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||  
 43 ATAAGCTTGGGAGCAGACCAAGCAGCTCTTGACCTCAATTAGCAGAGGA 92  
 54 uLeuSerGluIleSerLeuValLysHisIleTyrGluGlyLeuValGlnG 71  
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 93 TAATGTGCGCATCAAAAATGATGTGACACAAATGTTAGAGGGATTTACAG 142  
 71 LuAsnAsnLeuSerGlyAsnIleGluProAlaLeuAlaGluAspTyrSer 87  
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 143 GAGATCCCTAATACAGGGGAAATAAACCGGACCTTGCAAAAGGGTGGAT 192  
 88 LeuSerSerAspGlyLeuThrTyrThrPheLysLeuLysSerAlaPhe.. 103  
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 193 ATTTCTTCTGATGAACAGTTTACATTTTAACTTACAGAGAGAGAGAGATC 242  
 104 TrpSerAsnGlyAspProLeuThrAlaGluAspPheIleGluSerTrpL 120  
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 243 TTGGAGTACGGAGTTCATCTACTGACAGAGGAATTAGAAAAATCTTATC 292  
 120 ySlnValAlaThrGlnGluValSerGlyIleTyrAla...PheAlaLeu 135  
 : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 293 TTAGAAATTTAAATAAGAACTGGCTCAAACTACGTTGAATGGTTAAA 342

Sequence 1587 BP; 648 A; 251 C; 259 G; 429 T; 0 other;

## alignment\_scores:

Quality: 505.50 Length: 485  
 Ratio: 1.620 Gaps: 15  
 Percent Similarity: 64.330 Percent Identity: 28.041

## alignment\_block:

US-09-824-567-2 x AAX61763

Align seg 1/1 to: AAX61763 from: 1 to: 1587

```

38 IleAsnIleArgAspGluProArgSerLeuAspProArgGlnValArgLeu 54
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
94 ATAAGCTTGGAGCAGACGCAAGCAGCTCTGCACCTCAATAGCAGAGA 143
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
54 uLeuSerGluIleSerLeuValLysHisIleTyrGluGlyLeuValGlnG 71
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
144 TAATGTCGCATCAAAATGATTGACACAAATGTTAGAGGATTTGATACAG 193
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
71 luAsnAsnLeuSerGlyAsnIleGluProAlaLeuAlaGluAspTyrSer 87
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
194 GAGATCTTAATACAGGGGGAATAAACCAGGACTTGCAAAAGGGTGGAT 243
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
88 LeuSerSerAspGlyLeuThrTyrThrPheLysLeuLysSerAlaphe... 103
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
244 ATTCTCTCTGTGGAACAGCTTACACATTTAACTAAGAGAAAAATCAC 293
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
104 .TrpSerAsnGlyAspProLeuThrAlaGluAspPheIleGluSerTrpL 120
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
294 TTGGAGTACGAGCTGCATCACTGAGAGGAATTTAGAAATCTTATC 343
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
120 ySglnValAlaThrGlnGluValSerGlyIleTyrAla...PheAlaLeu 135
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
344 TTAGAATTTTAATAAAGAACTGGCTCAAGTAGCTTGAATGGTAAA 393
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
136 AspProIleLysAsnValArgLysIleGlnGluGlyHisLeuSerIleAs 152
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
394 TCGTAAITTAATAATGTCAAAATATTTTGATGGACAAGTGACTGACTC 443
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
152 pHisPheGlyValHisSerProAsnGluSerThrLeuValValThrLeuG 169
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
444 TGAAGTTGGAATAGACGGATTGATGAAAAACATTAGAAATAACACTGG 493
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
169 luSerProThrSerHisPheLeuLysLeuAla...Leu 181
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
494 AATCACCAAAACCTTATTTATTGATATGTTAGTACCAATCATTTAT 543
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
182 ProValPhePheProValHisLysSerGlnArgThrLeuGlnSerLysSe 198
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
544 CCAGTA.....CCAGTTCATGTACCGAAAAGTATGCACAAACTGGAC 587
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
198 rLeuPro.....IleAsnSerGlyAlaPheTyrProLysAsnIleLeu 212
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
588 AAGCCCGGAAACATGTCGACAGTGGCTTTTAAATTAAGAAAGAA 637
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
212 ySglnLysGlnThrIleLysLeuSerLysAsnProHisTyrTrpAsnGln 228
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
638 TTCCTAACGAAATAATGCTTTTGAATAAATAACAAATACTACGACTCA 687
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
229 SerGlnValGluThrLysThrIleThrIleHisPheIleProAspAlaAs 245
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
688 AATGANGTAGAATTAAGAGAGATTACATTTTACACAAATAATGACAGCTC 737
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
245 nThrAlaAlaLysLeuPheAsnGlnGlyLysLeuAsnTrpGlnGlyProp 262
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
738 AACAGCTATAAATGATGAAATGAAGAGCTAGAT.....GCAA 778
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
262 rofTrpGlyGluArgIleProGlnGlnThrLeuSerAsnLeuGlnSerLys 278
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
779 TTTTGTGTTCC...ATCCCCCAATCTAATCAAAATCTAAATTAAGA 825
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
279 GlyHisLeuHisSerPheAspValAlaGlyThrSerTrpLeuThrPheAs 295
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

...  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
826 AGGACTATTACTCATCAGCTGTTAATGCCATATACTTTTACGCGTTCAA 875
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
295 nIleAsnLysPheProLeuAsnAsnMetLysLeuArgGluAlaLeuAlaS 312
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
876 TACACACATCAAAACCACCTTGACACAGTTAAATTTAGAAAGCCTTAATC 925
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
312 erAlaLeuAspLysGluAlaLeuValSerThrIlePheLeuGlyArgAla 328
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
926 TTGCTATTGACAGAGAAACGCTTACATATAAAGTT...CTTGACACGGG 972
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
329 LysThrAlaAspHisLeuLeuProThrAsnIleHisSerTyrProGluH 345
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
973 ACTACCCCTACAAAGAGCAACTCCCAACTTTAGTTCATAT..... 1014
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
345 sGlnLysGlnGluMetAlaGlnArgGlnAlaTyrAlaLys.....LysL 360
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1015 .....TCATTATGCAAAAGTTTAGAAT 1036
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
360 euPheLysGluAlaLeuGluGluLeuGlnIleThrAlaLysAspLeuGlu 376
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1037 TATTTAATCTGAAATTTGCAAAACCCCTTCTAGCTGAAGCTGGATATCCT 1086
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
377 HisLeuAsnLeuIlePheProVal.....SerSerSe 387
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1087 AATGCAAT...GGATTCCAATTTTAAATTTAAATACAAATACAAACGA 1133
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
387 xAlaSerSerLeuLeuValGlnLeuIleArgGluGlnTrpLysGluSerL 404
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1134 AGCAAAATAAAAAATTTGTGAATTTATTCAAAACCAATGCAAAAAAAT 1183
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
404 euGlyPheAlaIleProIleValGlyLysGluPheAlaLeuLeuGlnAla 420
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1184 TAATATTGATGGAGACTTGAACAGCAAGATGGACAACATACTATAAC 1233
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
421 AspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGlyTrpPheAlaAs 437
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1234 ACTAAGCAAAATGGAATTTATGAAATAGCAAGAGCAGGATGGATAGGCGA 1283
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
437 pPheAlaAspProMetAlaPheLeuThrIlePheAlaTyrProSerGlyV 454
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1284 TTATGCTGATCCCTTGACATTTTAAAGCATTTTCA.....CAAGSAT 1327
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
454 alProProTyrAla.....IleAsnHisLysAspPheLeuGluIle 467
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1328 ACACAAATTCATCTCATTAATTAATCAACCCAGATACAAAGAACTT 1377
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
468 LeuGlnAsnIleGluGlnGluAspHisGlnLysArgSerGluLeuVa 484
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1378 ATAAAGAAATCGACCTTGACCTTGATCCCAATAAAAAAGACAAACATTTT 1427
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
484 lSerGlnAla.....SerLeuTyrLeuGluThrPheHisIleIleGluPro 500
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1428 AGACAAAGCAGAGAGATAATTTGAAAAAGATTTTCCCAATAGCACCAA 1477
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
500 leuTyr 501
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1478 TATAC 1482
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

seq\_name: /SIDS5/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AAAX20248

## seq\_documentation\_block:

ID AAX20248 standard; DNA; 910715 BP.

AC AAX20248;

XX 04-MAY-1999 (first entry)

XX Borrelia burgdorferi polynucleotide sequence #1.

XX Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;

XX epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;

XX infection; diagnosis; characterisation; detection; ds.





```

421 AspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGlyTrpPheAlaAs 437
572850 ACTAAGGCAAAATGGAATTAAGAAATAGCAAGAGCAGATGATAGCGGA 572801
437 pPheAlaAspProMetAlaPheLeuThrIlePheAlaTyProSerGlyV 454
572800 TTATGCTGATCCTTGATGATTTTAAAGCATATTCACA.....CAAGGAT 572757
454 alProTyroAla.....IleAsnHisLysAspPheLeuGluIle 467
572756 ACACACAATTCATCTCATATTAATTAATCAACCCAGATATCAACGAAT 572707
468 LeuGlnAsnIleGluGlnGluGlnAspHisGlnLysArgSerGluLeuVa 484
572706 ATAAGAAATCCGACCTTGAGTTGATCCATAAATAAAGACAGACATTT 572657
484 lserGlnAla...SerLeuTyLeuGluThrPheHisIleIleGluProI 500
572656 AAGACAGACAGAGAGATATTAATTTGAAAAAGATTTTCCAATAGCACCA 572607
500 leTyro 501
572606 TATAC 572602

```

seq\_name: /STD5/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AAAX20020

seq\_documentation\_block:

ID AAAX20020 standard; DNA; 1659 BP.

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

XX AC AAAX20020;

```

CC activity.
XX
SQ Sequence 1659 BP; 582 A; 305 C; 359 G; 413 T; 0 other;

alignment_scores:
Quality: 455.00 Length: 539
Ratio: 1.346 Gaps: 15
Percent Similarity: 62.709 Percent Identity: 25.232
alignment_block:
US-09-824-567-2 x AAAX20020
Align seg 1/1 to: AAAX20020 from: 1 to: 1659

4 IleSerValGlyIleCysIleThrIleLeuLeuSerLeuSerValVal 20
16 ATGAATTTGGGGAAAAAGTAGTAGTTTATTGATGCAACAGGGTTCTTT 65
20 uGlnGlyCys.....LysGluSerSerHis 29
66 AGCCGCGATGTGGCGGAACCAAGAGCGGCGAGAAAGTAGATTCGGAA 115
29 erSerThrSerArgGlyGluLeuAlaIleAsnIleArgAspGluProArg 45
116 ATTAGCAGCTGAACAAAAAATCAGTATTAGTTTACCTGCACCAATCTCA 165
46 SerLeuAspProArgGlnValArgLeuLeuSerGluIleSerLeuVal 62
166 ACATTTGGATACACACAAACACAGATAAAATACCTTTACATATGCACA 215
62 shisIleTyroGluGlyLeuValGlnGluAsnAsnLeuSerGlyAsnIle 79
216 ACATTTATTTGAAGGCTTTATCGGTTTGTATGATGATGATGATGATG 264
79 luProAlaLeuAlaGluAspTyroSerLeuSerSerAspGlyLeuThr 95
265 ..CCAGCTCTAGCTAAAGATGTCAAGATTAGTGACGATGGCGCAAGTAC 312
96 ThrPheLysLeuLysSerAlaPhe...TrpSerAsnGlyAspProLeu 111
313 CACTTTACCTTGGGAGGGGATTAAAGTGGAGCAACGGCGGCAATCAC 362
111 rAlaGluAspPheIleGluSerTrpLysGlnValAlaThrGlnGluVal 128
363 GGCCCAAGATTTGTTTATCTTGGAAAAAATCTGGTGACACAGCGACGA 412
128 erGlyIleTyroAlaPheAlaLeuAsnProIleLysAsnValArgLys 144
413 TTGGACCGAATGCTTATTACTAGACAGTGTAAAAATAGTTTGAATA 462
145 GlnGluGlyHisLeuSerIleAspHisPheGlyValHisSerProAsn 161
463 CGCAACGGTGAAGAACTCAGTCGATGAATAGGATTTTCAGCCCGCAATGA 512
161 uSerThrLeuValValThrLeuGluSerProThrSerHisPheLeuLys 178
513 CAAAGAAATTCATTTGTTGAATTAACACAGCCCAACCTTCCTCTTAC 562
178 eLeuAlaLeuProValPhePheProValHisLysSerGlnArgThrLe 194
563 TCGTTTCGATTTGTTTGGTTAGCGCCCAAAATCAAAATTTTTCGAGCG 612
195 GlnSerLysSerLeuPro.....IleAlaSerGlyAl 205
613 CAAGCAAGAAATTAAGCTTGGATAGTGAACATTTACTTTATAGCGGGC 662
205 aPheTyroProLysAsn.....IleLysGlnLysGlnTrpIleLysLe 220
663 ATTTACCTAGCAATTTGGATGGCCTTACACTTCTTACACTTGG...A 709
220 erLysAsnProHisTyroTyroAsnGlnSerGlnValGluThrLysThr 236

```

The present sequence represents a gene isolated from *Enterococcus faecalis*. The present invention describes genes, proteins and antigenic polypeptides isolated from *E. faecalis*. The proteins can be used in vaccines for preventing or attenuating an infection caused by a member of the *Enterococcus* genus in an animal. They can also be used for detecting *Enterococcus* antibodies in a sample. The nucleotide sequences can be used for detecting *Enterococcus* nucleic acids. Products from the present invention can also be used for screening compounds to identify agonists and antagonists of *E. faecalis* protein

Claim 1: Page 89-90; 301pp: English.

WPI: 1999-070095/06.  
P-PSDB: AAY000030.

New isolated *Enterococcus faecalis* polynucleotides - used to develop products for the detection of *Enterococcus* and for use in vaccines for prevention or attenuation of *Enterococcus* infection

04-MAY-1998; 98WO-080959.

14-NOV-1997; 97US-0066009.

06-MAY-1997; 97US-0044031.

16-MAY-1997; 97US-0046655.

(HUMA-) HUMAN GENOME SCI INC.

W09850554-A2.

12-NOV-1998.

20-APR-1999 (first entry)

*Enterococcus faecalis* gene EF012.

*Enterococcus faecalis*; infection; vaccine; immune response; diagnosis; detection; attenuation; antigenic; ss.

*Enterococcus faecalis*.

W09850554-A2.

12-NOV-1998.

04-MAY-1998; 98WO-080959.

14-NOV-1997; 97US-0066009.

06-MAY-1997; 97US-0044031.

16-MAY-1997; 97US-0046655.

(HUMA-) HUMAN GENOME SCI INC.

W09850554-A2.

12-NOV-1998.

20-APR-1999 (first entry)

*Enterococcus faecalis* gene EF012.

*Enterococcus faecalis*; infection; vaccine; immune response; diagnosis; detection; attenuation; antigenic; ss.

*Enterococcus faecalis*.

W09850554-A2.

12-NOV-1998.

04-MAY-1998; 98WO-080959.

14-NOV-1997; 97US-0066009.

06-MAY-1997; 97US-0044031.

16-MAY-1997; 97US-0046655.

```

710 AAAAAAATCAGAAATACTATGATCGGATCAAGTCAAACTGGAAGAAGTT 759
237 ThrileHisPheIleProAspAlaAsnThrAlaAlaLysLeuPheAsnG1 253
      ::::: ||| :::: |||||::: ::::: |||||:::
760 GCGGTTAGCACATCAAGACAGATAACTGGGATTAACCTTATATCAAGT 809
      ::::: ||| :::: |||||::: ::::: |||||:::
253 nGlyLysLeuAsnTrpGlnGlyProProTrpGlyGluArgIleProGlnG 270
      ::::: ||| :::: |||||:::
810 GAATGAACACTAGACTTA.....GTTCCGANTTAACGGAC 841
270 luThrLeuSerAsnLeuGln...SerLysGlyHisLeuHisSerPheAsp 285
      ::::: ||| :::: |||||:::
842 AATATGTTCACAATATCAAGATCAAGGCTATGTCAGTCAATCCAGAT 891
286 ValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLys...PheProLe 301
      ||||| :::: |||||:::
892 GTGGCC...AACTACTCTTGTAGATTCAACAAAAAGAGAACCCCAT 938
301 wasAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGluA 318
      ||||| :::: |||||:::
939 AGCGAATGTTTATTACGAAAGCGATTGGCCAAAGCAATTGATAAGAG 988
318 laLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisLeu 334
      ||||| :::: |||||:::
989 CCTTAACACAAAGTCTTAAACAGATGGGTCAAAACCCCTTAACGGATT 1038
335 LeuProThrAsnIleHisSerTrpProGluHisGlnLysGlnGluMetAl 351
      ::::: ||| :::: |||||:::
1039 ATTCCAAGTAACTTTATGCGAATCCAGAAACGGATGGAAGATTCCGAGC 1088
351 aGlnArgGlnAlaTyrrAlaLysLysLeuPheLysGluAlaLeuGluL 368
      ||||| :::: |||||:::
1089 TTACAGTGGCGAATTTGAAATAATGACGTCAAAAGAGCTCAAGCTGAAT 1138
368 euGlnIleThrAlaLysAspLeu...GluHisLeuAsnLeu...IlePhe 382
      ::::: ||| :::: |||||:::
1139 GGACGAAAGCCCAAGCGGATGCTGTAAAGAAAGTGAACCTTCAATGCTG 1188
383 ProValSerSerAlaSerLeuLeuValGlnLeuIleArgGluG1 399
      ::::: ||| :::: |||||:::
1189 GCGGACACACAGATCAAGCAAGCAATGCTGAATATGTTCAAAAGTCA 1238
399 nTrpLysGluSerLeu...GlyPheAlaIleProIleValGlyLysGluP 415
      ::::: ||| :::: |||||:::
1239 GTTCAGCAAAATGCGCGAGTTTGAATAATACCAATTTTCATCGCAACCA 1288
415 heAlaLeuLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThr 431
      ::::: ||| :::: |||||:::
1289 GTAATAATGTGAACCAATCGCGACGTGAAAAAATATGAGTTGTCTCT 1338
432 GlyGlyTrpPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePh 448
      ::::: ||| :::: |||||:::
1339 TCAGATGGATTCGCGGCAAGTGTAGTACCTTACTTTAACTTATA 1388
448 eAlaTyrrProSerGlyValProProTyrrAlaIleAsnHisLys..... 462
      ::::: ||| :::: |||||:::
1389 TGCAGGAGATCAAGTTAC...AATTCGGAATTTATCAATATGCAAAAT 1435
463 .....AspPheLeuGluIleLeuAsnIleGluGlnGluAspHis 477
      ::::: ||| :::: |||||:::
1436 ACGACCAATTTGTAGAGAGCGACCGACGATTAATGCCAATATCCAGAG 1485
478 GlnLysArgSerGluLeuValSerGlnAlaSerLeuTyrrLeuGluThrPh 494
      ::::: ||| :::: |||||:::
1486 AAACAGTTTCAGATACAAAGAGAGGAGACATCTTGTGAACCAAGA 1535
494 eHisIleIleGluProIleTyrrHisAspAlaPheGlnPheAlaMetAsnL 511
      ::::: ||| :::: |||||:::
1536 TGCTGCCAAGTACCGCTGTATCAAGTGCCTCAATTTATCAATCAATC 1585
511 ysLysLeuSerAsnLeu 516
      ||||| :::: |||||:::
1586 CTAATTTGAAGGCATT 1602

```

seq\_name: /SIDS5/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.AAX13087

seq\_documentation\_block:

ID AAX13087 Standard; DNA; 11739 BP.

XX AAX13087.

AC 19-MAR-1999 (first entry)

DT Enterococcus faecalis genome contig SEQ ID NO:150.

DE Enterococcus faecalis; contig: detection; Enterococcal infection;

XX Enterococcus faecalis; contig: detection; Enterococcal infection;

KW vaccine; attenuation; computer readable medium; ds.

XX Enterococcus faecalis.

OS W09850555-A2.

XX 12-NOV-1998.

XX 04-MAY-1998; 98WO-080985.

PF 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Barash SC, Dillon PJ, Kunsch CA;

XX WPI; 1999-045171/04.

XX New isolated Enterococcus faecalis polynucleotides and polypeptides

PT - used to develop products for the detection of Enterococcus and for

PT use in vaccines for prevention or attenuation of Enterococcus

PT infection.

XX Claim 1; Page 862-868; 2084pp; English.

XX A computer readable medium has been developed which has recorded on it

CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.

CC AAX12938 to AAX13919 represent these nucleotide sequences which are

CC primary nucleotide sequences, also known as contigs. The computer-based

CC system can identify fragments of the Enterococcus faecalis genome with

CC commercial importance. The products can be used to detect the presence

CC of Enterococcus faecalis in samples. They can also be used for

CC diagnosing Enterococcal infection in an animal and monitoring

CC progression of disease, and for identifying agents which can be used to

CC modulate the growth or pathogenicity of Enterococcus faecalis, or

CC another related organism, in vivo or in vitro. In particular the

CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences

CC can be used in vaccines to prevent or attenuate an Enterococcal

CC infection.

XX Sequence 11739 BP; 3917 A; 2161 C; 2537 G; 3118 T; 6 other;

alignment\_scores:

Quality: 455.00 Length: 539

Ratio: 1.346 Gaps: 15

Percent Similarity: 62.709 Percent Identity: 25.232

alignment\_block:

US-09-824-567-2 x AAX13087

Align seg 1/1 to: AAX13087 from: 1 to: 11739

4 lleSerValGlyIleCysIleThrIleLeuLeuSerLeuSerValValle 20

2994 ATGAAATGGGAAAAAAGTAGTAGTTGTTCACACAGGGTTCTTT 3043

20 uGlnGlyCys.....LysGluSerSerHis 29

```

3044 AGCCGCAATGTCGCGGAACCAAGAGCGCGCAGAAAGTATCGGGAA 3093
      ||| ||| ||| ||| |||
29  ex-SerThrSerArgGlyGluLeuAlaIleAsnIleArgAspGluProArg 45
      :: ::::::::::::::::::::
3094 ATTAGACGCTGAACAAAATACAGTATTATGATTCACCTGCACCAATCTCA 3143
46  SerLeuAspProArgGlnValArgLeuLeuSerGluIleSerLeuVally 62
      ::::::::::|||:::
3144 ACATTGGATCAACACACACACACATAAAATACCTTTACAATGGCACA 3193
62  shisIleTyGluGlyLeuValcInGluAsnLeuSerGlyAsnIle6 79
      ::::::::::|||:::
3194 ACATTATTTGAAGGCTTTATCGTTGATGATGATAGTGCACCGGTG. 3242
79  luProAlaLeuAlaGluAspTySerSerLeuSerSerAspGlyLeuThrtyr 95
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3243 ..CCAGCTCTAGCTAAGATGTCAGAGATTAGTACGATGGCGCAAGTAC 3290
96  ThrPheLysLeuLysSerAlaPhe...TipSerAsnGlyAspProLeuth 111
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3291 CACTTTACCTTGGGGAGGGATTAAGTGGAGACACGGCGGCAATCAC 3340
111 rAlaGluAspPheIleGluSerTrpLysGlnValAlaThrGlnGluVal 128
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3341 GGCCCAAGATTGTGTTATCTTGGAATAAACTGGTGACACACCGCAGCA 3390
128 erGlyIleTyAlaPheAlaLeuAsnProIleLysValArgLysIle 144
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3391 TTGGACCGAATGCCATTTACTAGACAGTGTAAATAAGTTTGAATA 3440
145 GlnGluGlyHisLeuSerIleAspHisPheGlyValHisSerProAsnG 161
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3441 CGCAACGGTGAAGATCAGTCGATGATAGGATTTAGCCCGCAATGA 3490
161 uSerThrLeuValThrLeuGluSerProThrSerHisPheLeuLysL 178
      ::::::::::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3491 CAAGCAATTCATGTGTAATTAACAGCGCCCAACCTTCCTTCTTACGAG 3540
178 euLeuAlaLeuProValPhePheProValHisLysSerGlnArgThrLeu 194
      ::::::::::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3541 TCGTTTCGATGCTTGGTTAGCCCAACAAATCAAAATTTGTGGAAGCG 3590
195 GlnSerLysSerLeuPro.....IleAlaSerGlyAl 205
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3591 CAAGCAAGATTAACGCTTGGATGAGCAATTTACTTTATACCGGCC 3640
205 aPheTyProLysAsn.....IleLysGlnLysGlnTrpIleLysLeu 220
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3641 ATTTAGCTAGCCCAATTTGGATGCGGATTCAGATCTGG...ACATTGA 3687
220 erLysAsnProHisTyTyTrAsnGlnSerGlnValGluThrLysThrIle 236
      ::::::::::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3688 AAAAAATCCAGATACTATGATGCGGATCAAGTGAACCTGGAAGAGTT 3737
237 ThrIleHisPheIleProAspAlaSerThrAlaAlaLysLeuPheAsnG 253
      ::::::::::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3738 GCGGTAGCACATCAAGAAATATACTGGGATTAACCTTATATCAAGT 3787
253 nGlyLysLeuAsnTrpGlnGlyProTrpGlyGluArgIleProGlnG 270
      ::::::::::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3788 GAATGAACATAGACTTA.....GTCCGATTAAACGGAC 3819
270 luThrLeuSerAsnLeuGln...SerLysGlyHisLeuHisSerPheAsp 285
      ::::::::::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3820 AATATGTTCAACAATATCAAGATGATCAGGCTATGTCAGTATCCAGAT 3869
286 ValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLys...PheProLe 301
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3870 GTGGCC...AAGTACTTCTAGATTCAACAAAAAAGAGGACCCATT 3916
301 uAsnAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGlu 318
      | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

3917 AGCGAATGTTCAATTACGAAAGCGATTGGCCAAAGCAATGTATAAGAAG 3966
318 laLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisLeu 334
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3967 CCTTAACACAAAGTCTCTTAACGATGGGTCAAAACCCCTTAACGGATTG 4016
335 LeuProThrAsnIleHisSerTyProGluHisGlnLysGlnGluMetAl 351
      ::::::::::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4017 ATTCCAAGTAACACTTTATGGCAATCCAGAAACGGATGAAGATTCCGAGC 4066
351 aGlnArgGlnAlaLysLysLeuPheLysGluAlaLeuGluGluL 368
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4067 TTACAGTGGCAATATTGAAATACGCTCAAAAGGCTCAAGCTCAAT 4116
368 euGlnIleThrAlaLysAspLeu...GluHisLeuAsnLeu...IlePhe 382
      :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4117 GGACGAAAGCCCAAGCGGATGTCGGTAAAAAGTGAACCTTTCAATTGCTG 4166
383 ProValSerSerAlaSerSerLeuLeuValGlnLeuIleArgGluG 399
      ::::::::::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4167 GCGGAGACACAGATCAAGGAAAGCAATTTGCTGAATATGTTCAAAGTCA 4216
399 nTrpLysGluSerLeu...GlyPheAlaIleProIleValGlyLysGluP 415
      ::::::::::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4217 GTTCAAGAAATCTCCAGGTTTAGAAATACCATTTTCATCGCAACCA 4266
415 heAlaLeuLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThr 431
      :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4267 GTAATAATGTGAACCAATCGCGACGCTGAAAAAATATGAGTTGCTCTT 4316
432 GlyGlyTrpPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePh 448
      ::::::::::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4317 TCAGATGAGTTCGCGCAGTAGTGAATTTAGACTCTTACTTTAACTTATA 4366
448 eAlaTyProSerGlyValProProTyAlaIleAsnHisLys..... 462
      :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4367 TGCAGGAGATCAAGTTAC...AATTACGCAATTTATCAATGCCAAAT 4413
463 .....AspPheLeuGluIleLeuGlnAsnIleGluGlnGlnAspHis 477
      ::::::::::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4414 ACAGCAATTTGGTAGAGAGCGCAGCAAGATTAATGCCAATATCCAGAG 4463
478 GlnLysArgSerGluLeuValSerGlnAlaSerLeuTyTrLeuGluThrPh 494
      ::::::::::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4464 AAACAGTTTGCAGATACAAAGACCGGAGACATCTTGTGTAACCAAGA 4513
494 eHisIleIleGluProIleTyHisAspAlaPheGlnPheAlaMetAsnL 511
      ::::::::::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4514 TGTGTCACCAAGTACCGCTGTATCAAAAGTCCCTCAATTAATCAATC 4563
511 yLysLysLeuSerAsnLeu 516
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4564 CTAAATGGAAGGCATT 4580

```

seq\_name: /SID55/gcdata/geneseq/geneseq-emb1/NA1999.DAT.AAX20021

seq\_documentation\_block:

```

ID AAX20021 standard; DNA; 1585 BP.
XX
AC AAX20021;
XX
DT 20-APR-1999 (first entry)
XX
DE Enterococcus faecalis EF012 gene fragment.
XX
KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KW detection; attenuation; antigenic; ss.
XX
OS Enterococcus faecalis.
XX
PN WO9850554-A2.
XX
PD 12-NOV-1998.

```

XX PF 04-MAY-1998; 98WO-US08959.  
 XX PR 14-NOV-1997; 97US-0066009.  
 XX PR 06-MAY-1997; 97US-0044031.  
 XX PR 16-MAY-1997; 97US-0046655.  
 XX PA (HUMA-) HUMAN GENOME SCI INC.  
 XX PI Bailey C, Choi GH, Bromocky J A, Kunsch CA;  
 XX DR NPI: 1999-070095/06.  
 XX DR P-PSDB; AAY00031.  
 XX PT New isolated Enterococcus faecalis polynucleotides - used to develop  
 PT products for the detection of Enterococcus and for use in vaccines  
 PT for prevention or attenuation of Enterococcus infection  
 XX PS Claim 1; Page 90-91; 301pp; English.  
 XX CC The present sequence encodes an antigenic polypeptide fragment  
 CC isolated from Enterococcus faecalis. The present invention describes  
 CC genes, proteins and antigenic polypeptides isolated from E. faecalis.  
 CC The proteins can be used in vaccines for preventing or attenuating an  
 CC infection caused by a member of the Enterococcus genus in an animal.  
 CC They can also be used for detecting Enterococcus antibodies in a sample.  
 CC The nucleotide sequences can be used for detecting Enterococcus nucleic  
 CC acids. Products from the present invention can also be used for  
 CC screening compounds to identify agonists and antagonists of E. faecalis  
 CC protein activity.  
 XX CC  
 SQ Sequence 1585 BP; 559 A; 296 C; 336 G; 394 T; 0 other;

alignment\_scores:  
 Quality: 452.50 Length: 519  
 Ratio: 1.380 Gaps: 17  
 Percent Similarity: 63.198 Percent Identity: 26.590  
 alignment\_block:  
 US-09-824-567-2 x AAX20021 ..  
 Align seg 1/1 to: AAX20021 from: 1 to: 1585

22 GlyCysLysGluSerSerHisSerThrSerArgGlyGluLeuAla... 37  
 8 GGAACCAAGAAAGCCGACAGAAAGTACATTCG...GGAAATTTACACC 54  
 38 .....ileAsnileArgAspGluPro.....ArgSerLeuAspP 49  
 55 TGAACAAAATACGATTAGTTAGTTCACCTGCACCAATCTCAACATTGGATA 104  
 49 roArgGlnValArgLeuLeuSerGluileSerLeuValLysHisIleTy 65  
 105 CAACACAAACACAGATAAAATACCTTTACATGCGACACACATTAATT 154  
 66 GluGlyLeuValGlnGluAsnLeuSerGlyAsnileGluProAlaLe 82  
 155 GAAGGCGCTTATCGGTTTATGATGATGATGATGATGATGATGATGATGAT 201  
 82 uAlaGluAspTySerLeuSerSerAspGlyLeuThrThrPheLysL 99  
 202 AGCTAAGATGTCAGATGATGATGATGATGATGATGATGATGATGATGAT 251  
 99 eulysSerAlaPhe...TrpSerAsnGlyAspProLeuThrAlaGluAsp 114  
 252 TGGCGGGGGGATTAAGTGGAGCAACGGCGCAATACACGCCCAAGAT 301  
 115 PheileGluSerTyPlyGlnValAlaThrGlnGluValSerGlyIleTy 131  
 302 TTTGTTTATTTCTGGAAAAACCTGGTACACACACGACGATGGACCGAA 351  
 131 rAlaPheAlaLeuAsnProLysAsnValArgLysIleGlnGlyH 148

352 TGCCTATTACTAGACAGTCTTAAATAATAGTTTGAATAGCAAGGTG 401  
 148 isLeuSerileAspHisPheGlyValHisSerProAsnGluSerThrLeu 164  
 402 AAAAGTCAGTCGATGAATAGGATTTTACGCCCGGATGACAAAGATTC 451  
 165 ValValThrLeuGluSerProThrSerHisPheLeuLysLeuAlaLe 181  
 452 ATTGTTGAATTAACACAGCCCACTTCCTTCTTACGACGTCGTTGAT 501  
 181 uProValPhePheProValHisLysSerGlnArgThrLeuGlnSerLys 198  
 502 TCGTTGTTAGCGCCCAAAATCAAAATTTTGTCAAGCCCAAGGCAAG 551  
 198 erLeuPro.....ileAlaSerGlyAlaPheTyPro 208  
 552 ATTACGCGCTTGGATGATGAACATTTACTTTATAGCGGCGCATTTACGCTA 601  
 209 LysAsn.....ileLysGlnLysGlnTrpIleLysLeuSerLysAsnPr 223  
 602 GCCAATTTGGGATCGGACTTCAGACTTGG...ACATGAAAAAAATCC 648  
 223 oHisTyTyAsnGlnSerGlnValGluThrLysThrIleThrIleHisP 240  
 649 AGAATATGATGATGCGGATCAAGTGAAGTGAAGTGGGTTAGCA 698  
 240 heileProAspAlaAsnThrAlaLysLeuPheAsnGlnGlyLysLeu 256  
 699 CAATCAAGAAAGATTAATCTGGGATTAATATATCAATGAATGAACATA 748  
 257 AsnTrpGlnGlyProTrpGlyGluArgIleProGlnLysLeuLeuSe 273  
 749 GACTTA.....GTTCCGCAATTAACGGACAAATATGTCA 780  
 273 rAsnLeuGln...SerLysGlyHisLeuHisSerPheAspValAlaGly 289  
 781 ACAATATCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 827  
 289 hrSerTrpLeuThrPheAsnIleAsnLys....PheProLeuAsnMet 304  
 828 ACTACTTCTTAGATTTTCAACAAAAGAGAGAACCCCATTTAGCAATGTT 877  
 305 LysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGluAlaLeuValSe 321  
 878 CATTTACGAAAGAGGATTTGCGCAAGCAATTTGATTAAGAGCCTTAACA 927  
 321 rThrIlePheLeuGlyArgAlaLysThrAlaAspHisLeuLeuProThrA 338  
 928 AGTGTCTTAACAGTGGGTCAAAACCCCTTAACGGATGATTTCAAGTA 977  
 338 snIleHisSerTyProGluHisGlnLysGlnGluMetAlaGlnArgGln 354  
 978 AACTTTATGCGAATCCAGAAACGGATGAGATTTCCGAGCTTACAGTGGC 1027  
 355 AlaTyAlaLysLysLeuPheLysGluAlaLeuGluGluLeuGlnIleTh 371  
 1028 GAATATTGTAAAAATGACGTCAAAAGCTCAAGTGAATGGAGAAAGC 1077  
 371 rAlaLysAspLeu...GluHisLeuAsnLeu...IlePheProValSerS 386  
 1078 CCAAGCGGATGTCGGTAAAAAAGTGAACCTTTTCATTTGCTGGCGCAGACA 1127  
 386 erSerAlaSerSerLeuLeuValGlnLeuIleArgGluGlnTrpLysGlu 402  
 1128 CAGATCAAGGAAACCAATTTGCTGAATATGTTCAAGTCAAGTTCAGTCAAGAA 1177  
 403 SerLeu...GlyPheAlaIleProIleValGlyLysGluPheAlaLeuLe 418  
 1178 AATCTGCGAGGTTTAGAAATTTACATTTTCATCGCAACCAAGTAATAATGT 1227  
 418 uGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGlyTrp 435

1228 GAACCAATCGCGAGTCGAAAAAATAATGAGTGTCTCTTCAGGATGGA 1277  
 435 heAlaAspPheAlaAspProMetAlaPheLeuThrIlePheAlaTyrPro 451  
 1278 TTTCGGCGAGTGTAGTGAATTAAGACTCTTACTTTACTTATATGCGAGGAA 1327  
 452 SerGlyValProTyrAlaIleAsnHisLys .....AspPh 464  
 1328 TCAGTTAC...AATACGSCAATTATCATTAATGCCAAATACGACCAAT 1374  
 464 eLeuGluLeuLeuGluAsnIleGluGluGluAsnHisGluLysArgS 481  
 1375 GTGACAGACGCGACGACGACGACGACGACGACGACGACGACGACGACGAC 1424  
 481 erGluLeuValSerGluAlaSerLeuTyrLeuGluThrPheHisIleIle 497  
 1425 CAGAATCAAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1474  
 498 GluProIleTyrHisAspAlaPheGluPheAlaMetAsnLysLysLeuSe 514  
 1475 GTACCGGTGTCAAGTGCCTCAATTAATCTATCAATCTCAATCTCAATCTCA 1524  
 514 rAsnLeu 516  
 1525 AGGCATT 1531

seq\_name: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AAx20253

seq\_documentation\_block:

ID AAX20253 standard; DNA; 26811 BP.

AC AAX20253;

DT 04-MAY-1999 (first entry)

DE Borrelia burgdorferi polynucleotide sequence #6.

KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;  
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;  
 KW infection; diagnosis; characterisation; detection; ds.

OS Borrelia burgdorferi.

PN W09858943-AL.

PD 30-DEC-1998.

PF 18-JUN-1998; 98WO-US12764.

PR 03-SEP-1997; 97US-0057483.

PR 20-JUN-1997; 97US-0050359.

PR 22-JUL-1997; 97US-0053344.

PR 22-JUL-1997; 97US-0053377.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (MEDI-) MEDIMMUNE INC.

PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;  
 PI White OR;

DR WPI; 1999-081217/07.

XX New isolated Borrelia burgdorferi nucleic acids - used to develop  
 XX products for the detection, diagnosis, characterisation, prevention  
 XX and therapy of infections, particularly Lyme disease

PS Claim 1; Page 851-867; 1128pp; English.

XX AAX20248 to AAX20402 represent polynucleotide sequences isolated from  
 CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for  
 CC the detection, diagnosis, characterisation, prevention and therapy of  
 CC Bb infections, e.g. Lyme disease. They can also be used for the  
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs

CC to a family of motile, spiral-shaped bacteria called Spirochetes.  
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and  
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as  
 CC Lyme disease.

SQ Sequence 26811 BP; 8978 A; 3456 C; 3596 G; 10780 T; 1 other;

alignment\_scores:

Quality: 451.50 Length: 501

Ratio: 1.360 Gaps: 15

Percent similarity: 66.267 Percent identity: 26.347

alignment\_block:

US-09-824-567-2 x AAX20253/rev ..

Align seg 1/1 to reverse of: AAX20253 from: 1 to: 26811

1 MetArgLysIleSerValGlyIleCysIleThrIleLeuSerLeuSe 17  
 14530 ATAAAAAGTTAAAGTTGTTATTTCTCAATTAAATTTACTTATTC 14481

17 rValValLeuGlnGlyCysLysSerHisSerThrSerArgG 34  
 14480 TTGTGTT .....AATGAAGTAATAGAACAAAT 14452

34 LyGluLeuAlaIleAsnIleArgAspGluProArgSerLeuAspProArg 50  
 14451 TGGTTTTTAAGCTAAATATGGAAGTGGAGCTGCTACTTAGATGCTCAA 14402

51 GlnValArgLeuLeuSerGluIleSerLeuValLysHisIleTyrGluI 67  
 14401 TTAATAAACGATACGTTGGATCGGAGTGTGAGCCAAATGTTCTTGG 14352

67 YLeuValGlnGluAsnAsnLeuSerGlyAsnIleGluProAlaLeuAlaG 84  
 14351 CATTTAGATGAGATCCAGGACTGGAGGATACAGACGGGACTTGCTA 14302

84 LuAspTyrSerLeuSerSerAspGlyLeuThrThrPheLysLeuLys 100  
 14301 AAAGTTGGGATATTTCTCATGACGGAGTAGTTATATACGTTTCATTTAAGA 14252

101 ...SerAlaPheTyrSerAsnGlyAspProLeuThrAlaGluAspPheI 116  
 14251 GATAATCTGTGTTGGAGTATGAGTTTCCATTACTGCCGAGAGATAAG 14203

116 eGluSerTyrLysGlnValAlaThrGlnGluValSerGlyIleTyrAlaP 133  
 14202 AAAATCTTATCTTAGAATTTTAGATAAAGAAACCGGCTCATCTTTGTTA 14153

133 heAlaLeu...AsnProIleLysAsnValArgLysIleGlnGluGlyHis 148  
 14152 ACATGATTAAGTCGGTTTATTAATAATGACAGAGAGTATTTTGACGGCAA 14103

149 LeuSerIleAspHisPheGlyValHisSerProAsnGluSerThrLeuVa 165  
 14102 GCAATGAGCTGAGCTTGAATTAAGCTCTTGATGAAAAAACTTTAGA 14053

165 IValThrLeuGluSerProThrSerHisPheLeuLysLeuAlaLeuP 182  
 14052 AATAACGCTAAATCTCCAAAGCCATATTTCTTGATATGTTAGTACATC 14003

182 roValPhePheProVal.....HisLysSerGlnArgThrLeuGlnSer 196  
 14002 AAACATTTATCTCTGACCAATGACGCTTATTGAAAGATATGGCAAGG 13953

197 LysSerLeuPro.....IleAlaSerGlyAlaPheTyrProLysAs 210  
 13952 TGGACAGACCCGAGAAATATGTTGTTAGTGGTCTTCAAAATTAATC 13903

210 nIleLysGlnLysGlnTyrIleLysLeuSerLysAsnProHisTyrTyr 227  
 13902 TAGAGTTTAAATGAAAGGTTGTTCTTGAAGAGATAATAATATTATA 13853

```

227 snGlnSerGlnValGluThrLysThrLysThrLysPheIleProAsp 243
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13852 ATTCGAAGATGTTGTTGACAGTATT...ATATTTTTCACAGAT 13806
244 AlaAsn...ThrAlaLysLeuPheAsnGlnGlyLysLeuAsnTrpGI 259
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13805 AATGACATTACAGCTTATAATATGAT...TTAAATGATGA 13768
259 nGlyProProTrpGlyGluArgIleProGlnGluThrLeuSerAsnLeuG 276
GCTGGATGCAATTTTAAAGATGTTCCACACAGATTGCTTAAGGATCTTA 13718
276 lnSerLysGlyHisLeuHisSerPheAspValAlaGlyThrSerTrpLeu 292
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13717 AGCTTAGGACGACGATTATTATCAATGGGTATTAAATCAACCTCTTTTAT 13668
293 ThrPheAsnIleAsnLysPheProLeuAsnAsnMetLysLeuArgGluAl 309
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13667 TCTTTGAACATCAAGTAAGAACCCCTTGCAATGTTAAAGTTAGAAAGC 13618
309 aLeuLaSerAlaLeuAspLysGluAlaLeuValSerThrIlePheLeuG 326
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13617 GCTGTCTTTGCTATTGATAGAAAAACCTTACACAGAGCGCTT... 13572
326 lyArgAlaLysThrAlaAspHisLeuLeuProThrAsnIleHisSerTyr 342
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13571 .....AATGATAGTTCTATTCTCTACAAGA...AGAGCAACT 13539
343 ProGluHis.....GlnLysGlnGluMetAlaGlnArgGlnAl 355
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13538 CCAGATTATTATGTTACTCTTATAAAGCAATTTGAGCTTATTGATGC 13489
355 a...TyrAlaLysLysLeuPheLysGluAlaLeuGluGluLeuIleT 371
13488 TGAATGCGAAGAAGCTTTGCGAGATCGAGATATCCC.....A 13448
371 hrAlaLysAspLeuGluHisLeuAsnLeuIlePheProValSerSer 387
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13447 ATGGTAATAATTTTCCTTTATTAAAGTAAAGTACATAAGTGATAGC 13398
388 AlaSerSerLeuLeuValGlnLeuIleArgGluGlnTrpLysGluSerLe 404
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13397 CAGAGAAA...ATTGCTGAATTTATTCAAAATCAGTGAAGAAAAACTT 13351
404 uGlyPheAlaIleProIleValGlyLysGluPheAlaLeuGlnAlaA 421
13350 AAATATTAAATGATCAGCTTGCAATGAAGATGCTCAACATATAAATA 13301
421 spLeuSerSerGlyAsnPheSerLeuAlaThrGlyGlyTrpPheAlaAsp 437
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13300 GTAGCTAAATGGTAATTAATGAATAATAAGATCAGGATGGTCAGGAGAT 13251
438 PheAlaAspProMetAlaPheLeuThrIlePhe...AlaTyrProSerGI 453
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13250 TATCTGATCTATGACGTTCTTAGCATCTTTCAAACTGAAATACATAC 13201
453 yValProProTyrAlaIleAsnHisLysAspPheLeuGluIleLeuGlnA 470
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13200 TTTTTCATCTTGTGATATCAAAATCTGAATATGATGACTTTTAATAA 13151
470 snIleGluGlnGluAspHisGlnLysArgSerGluLeuValSerGln 486
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13150 AATCAGATATGAGAGAGATATTTTAAAGACAGAGAAATCTTAAAAA 13101
487 Ala 487
|||
13100 GCA 13098

```

```

seq_name: /SID85/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AA13139
seq_documentation_block:
ID AAX13139 standard; DNA; 10996 BP.

```

```

XX AAX13139;
XX
XX 19-MAR-1999 (first entry)
XX
XX Enterococcus faecalis genome contig SEQ ID NO:202.
XX
XX Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
XX
XX Enterococcus faecalis.
XX
XX W09850555-A2.
XX
XX 12-NOV-1998.
XX
XX 04-MAY-1998; 98WO-US08985.
XX
XX 14-NOV-1997; 97US-0066009.
XX
XX 06-MAY-1997; 97US-0044031.
XX
XX 16-MAY-1997; 97US-0046655.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Barash SC, Dillon PJ, Kunsch CA;
XX
XX WPI; 1999-045171/04.
XX
XX New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
XX
XX Claim 1; Page 1065-1071; 2084pp; English.
XX
XX A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAX12938 to AAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
XX
XX Sequence 10996 BP; 3633 A; 1848 C; 2288 G; 3223 T; 4 other;

alignment_scores:
Quality: 443.00 Length: 526
Ratio: 1.443 Gaps: 13
Percent Similarity: 58.365 Percent Identity: 26.046

alignment_block:
US-09-824-567-2 x AAX13139
..
Align seg 1/1 to: AAX13139 from: 1 to: 10996

30 SerThrSerArgGlyGluLeuAlaIleAsnIleArgAspGluProArgSe 46
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 ACTGCCGCTAAACAAGANTTTAAAGTAGTAGTTCAACAAGAAATGCCTTC 50
46 rLeuAspProArgGlnValArgLeuLeuSerGluIleSerLeuValLysH 63
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
51 TGCAGATTATCTTTAGGAACAGATACAAATTAGTTTACGGCACTGAATA 100
63 isIleTyrGluGlyLeuValGlnGluAsnAsnLeuSerGlyAsnIleGlu 79
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



```

101 ATGCGTATGAAGGAATTTATCGTTTAGATGATTAAGCAAGCCGCAACCA 150
80 ProAlaLeuAlaGluAspTyrSerLeuSerSerAspGlyLeuThrThr 96
151 CGAGGGCTAAAGAAAGAGTGAAGTCAGTCACTGATGCTGCTGACGTACAC 200
96 rPheLysLeuLys...SerAlaPheTrpSerAsnGlyAspProLeuThr 112
201 AGTGAACACTAAGAGAGCAAAATGGTCCATGGTGATCCTGTCACAG 250
112 laGluAspPheLeuSerTrpLysGlnValAlaThrGlnGluValSer 128
251 CTGCGCATATATTTTATGGCAACGTACAGCAGATCCTCAGACTGGC 300
129 GlyIleThrAlaPheAlaAsnProIleLysAsnValArgLysIleGl 145
301 GCTGAGTATGCTATTTCTTGAATATGTTGAACCGGTGACAGCATGT 350
145 nGluGlyHisLeuSerIleAspHisPheGlyValHisSerProAsnGlu 162
351 AGCTGGTAAGAAACCAAGTCTCTGAATTAGCATTAAGCAACGAGATT 400
162 erThrLeuValThrLeuGluSerProThrSerHisPheLeuLysLeu 178
401 ACAGCTAGCAATTAACCTGGCAAAACCAACACCATATTTGATATTTA 450
179 LeuAlaLeuProValPhePheProValHisLysSerGlnArgThrLeuGl 195
451 CTGGCTTTCCACTATTTTCCGCGCAACAT.....CAAGCAACCGPAGA 494
195 n.....SerLysSerLeuProIleAlaSerG 204
495 GAAATATGGCAAGATTACGCTCTCCAGTGAAGAAAGCACTCTATAATG 544
204 lylAlaPheTyrProLysAsnIleLys.....GlnLysGlnTrp 216
545 GTCCGTTTGTCTTCACTAATTTGAAGGGCTGCTCAGATACCAATTGG 594
217 IleLysLeuSerLysAsnProHisTyrTrpAsnGlnSerGlnValGluThr 233
595 ...ACTTTGGAAAAAATGAACACTATTGGGATTAAGACAAATGTCAAAAT 641
233 rlysthrIleThrIleHisPheIleProAspAlaAsnThrAlaAlaLysL 250
642 AGATAAAATCAATTTGATGTTGTCAAGAGGCGCCCAACCGCTTCAACT 591
250 euPheAsnGlnGlyLysLeuAsnTrpGlnGlyProTrpGlyGluArg 266
692 TATTCGAAGTGGCCAGGCGAGC.....GATGTTATT 723
267 IleProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSe 283
724 TTAAGCGCGCAATTTGCCAGCAAGATCGCAAGGATCCAGATTTGGTTAT 773
283 rPheAspValAlaGlyThrSerTrpLeuThrPheAsn.....IleAsnL 298
774 CGAAAAAGAGACGACGACCAAGCTATTGGAAATTTATCAGCCGAGATAAA 823
298 ysPheProLeuAsnMetLysLeuArgGluAlaLeuAlaSerAlaLeu 314
824 ATTCACCGTATAATTAATGTAATTTAGCCAAAGCAATTTTCAGCAGCGATT 873
315 AspLysGluAlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAl 331
874 GATCGCAACGCTATTAGTAGATAAATTTTAGCGAGCGCTCAGTAGTCGC 923
331 aAspHisLeuLeu.....ProThrAsnIleHis 341
924 TACCGGCTTAATCCAGAGGATGTCGTATTTCGCCGACAGATGACACAG 973
341 erTyrProGluHisGlnLysGlnGluMetAlaGlnArgGlnAlaTyrAla 357
974 ATTTTCTGATGAAAAATAAAAAAATTTTGAATATATAGCCGCAAGCA 1023

```

```

358 LysLysLeuPhePheLysGluAlaLeuGluGlnIleThrAlaLysAs 374
1024 AAAGAAATATTGGCAAAAGCAAGAAATTTGGGCATAACAACACTCAA 1073
374 pleuGluHisLeuAsnLeuIlePheProValSerSerSerAlaSerSerL 391
1074 AATGGAT.....ATTGTTGCCGATGATGCGATTCTACGAAA 1111
391 euLeuValGlnLeuIleArgGluGlnTrpLysGluSerLeu...GlyPhe 406
1112 AGTTAGCGGAATATTTCAAGGTACGTTTAAAGGATCTTTAGAGGCAATT 1161
407 AlaIleProIleValGlyLysGluPheAlaLeuLeuGlnAlaAspLeuSe 423
1162 GATGTACACAGTGAGCCCGTTCGCGTTTTCAGTTTCAATTCGAATTGATCGAGGTAG 1211
423 rSerGlyAsnPheSerLeuAlaThrGlyGlyTrpPheAlaAspPheAlaA 440
1212 TCGCGGCGATTTTGAGACAATCTTAGTGGTTGGCTGCGGATTACGCAG 1261
440 spPrometalAlaPheLeuThrIlePheAlaTyrProSerGlyValPro 456
1262 ATCCAGTAGTTTCTTAGATCTTTTGTACTGTGTAATAACTATAATATCGC 1311
457 TyrAlaIleAsnHisLysAspPheLeuGluIleLeuGln.....AsnIl 471
1312 GGGCGTTTTCAGTAAAGCTACGATGATGATTGAGGCTTCAGGTAC 1361
471 eGluGlnGluAspHisGlnLysArg...SerGluLeuValSerGlnA 487
1362 AAGAGATGCTTCAGATCCTGAAACACCTTCGGAGACATGGTTAAGGCTG 1411
487 laSerLeuTyrLeuGluThrPheHisIleLeuGluProIleTyrHisAsp 503
1412 AGAAGTTATTGATGGTGAAGAGACCGCTTTAGCACCACTTTATCAAAAG 1461
504 AlaPheGlnPheAlaMetAsnLysLysLeuSerAsnLeuGlyValSerPr 520
1462 GCGACTGCCCATCTACGCAGTAAAGAGTGAAGGCGTCTGTCGCCATGG 1511
520 oThrGlyVal...ValAspPheArgTyr 528
1512 TGTGTCGCGCACAAATATGATTATAAGTGG 1539

```

seq\_name: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.AAS46237

seq\_documentation\_block:

ID AAS46237 standard; DNA; 12790 BP.

AC AAS46237;

DT 18-DEC-2001 (first entry)

DE DNA encoding novel mar regulated protein (NIMR) #6.

KW mar regulated polypeptide; NIMR; microbial infection; antibacterial; ds.

OS Escherichia coli.

PN WO200170776-A2.

PD 27-SEP-2001.

PF 08-MAR-2001; 2001WO-US07478.

PR 10-MAR-2000; 2000US-188362P.

PA (TUFT ) TUFTS COLLEGE.

PI Levy SB, Barbosa TW, Aleksun MN;

DR WPI; 2001-602769/68.

DR P-PSDB; AAU29335.

Identifying compounds that modulate a newly identified mar regulated polypeptide activity, useful as antimicrobial compounds, involves contacting the polypeptide with a test compound -

PS Disclosure; Page 170-177; 526pp; English.  
vv

The invention relates to a method of identifying compounds that modulate a newly identified mar regulated (NIMR) polypeptide activity. The method comprises contacting an NIMR polypeptide with a test compound under interaction conditions, determining the ability of the compound to modulate the activity or expression of the polypeptide, and selecting the modulators. NIMR nucleic acids and polypeptides are used in the treatment of microbial infections, and in screening for modulators of NIMR expression and activity. These modulators can be used to reduce the infectivity of a microbe on a surface, and the virulence of a microbe in a subject suffering from an infection. AAS46232-AAS46278 represent *Escherichia coli* NIMR coding sequences of the invention.

SQ Sequence 12790 BP; 3387 A; 3116 C; 3135 G; 3152 T; 0 other:

alignment\_scores:

Quality:	438.50	Length:	515
Ratio:	1.410	Gaps:	12
Percent Similarity:	60.388	Percent Identity:	26.019

alignment\_block:

US-09-824-567-2 x AAS46237

Align seg 1/1 to: AAS46237 from: 1 to: 12790

4 IleSerValGlyIleCysIleThrIleLeuLeuSerLeuSerValValle 20  
1341 GTTTCAGTCACTGTTGGTGGCTGTGTGGTTCAGCAGCAATTTCTTCTCGTA 1390  
20 uGlnGlyCysLysGluSerHisSerSerThrSerArgGlyGluLeuA 37  
1391 TCGTCGAGAAGTTCGACGCGCAGCTACTGGCAGAGAAGCAGAGAGCTGG 1440  
37 lalleAsnIleArgAspGluProArgSerLeuAspProArgGlnValArg 53  
1441 TGGCCACACATTAAGATGAGCTCGCTCGCTGGATCCCGCTAAAGCCGCTG 1490  
54 LeuLeuSerGluIleSerLeuValLysHisIleTyrrGluGlyLeuValG 70  
1491 GGCTGCCAGAGATTCAGCTCATTTGCGGATCGTGTGTAAGTCTGTGGTAA 1540  
70 nGluAsnLeuSerGlyAsnIleGluProAlaLeuAlaGluAspTyrS 87  
1541 TCAGAACGAA...AAGGGGAGATTGTCCCGGCGCTTCGCACTCAGTGGGA 1587  
87 erLeuSerSerAspClyLeuThrTyrrPheLysLeuLys...SerAla 102  
1588 AA...AGTATGACACCGATCTCGACTTTTACCCTGGCGATAACGCA 1634  
103 PheTrpSerAsnGlyAspProLeuThrAlaGluAspPheIleGluSerTr 119  
1635 AAATGGCGGATGTCACACCGCTACGCGGCAAGATTTTCTTCACAGTGS 1684  
119 pLysGlnValAlaThrGlnGluValSerGlyIleTyrrAlaPhe.....A 134  
1685 GCAACGCTGTGTGGACCCAAAAACATTTGCCATTTGCATGTTTGCCTG 1734  
134 lalaLeuAsnProIleLysAsnValArgLysIleGlnGluGlyHisLeuSer 150  
1735 CGTGGCGGAATCAACAGCCGACAGCGGATATTGATGTGTAAGCTACG 1784  
151 IleAspHisPheGlyValHisSerProAsnGluSerThrLeuValValTh 167  
1785 CCTGACCAGCTGGCTGACCGAGTGTGCCATGCTCCATGTTGAAATCA 1834

```

2693 AGGAAATATTTACACCTTTAAACATCGCATATGACAAAGTCTGCGCC 2742
      ::::::::::::::::::::
470 snlledGluGlnGluGlnAspHisGlnLysArgSerGluLeuValSerGln 486
      ::::::::::::::::::::
2743 AGGCATCGACGAGAAATACCGTTAAAGCCGTAATGCGGATTACACAGCG 2792
      ::::::::::::::::::::
487 AlaSerLeuTyrLeuGluThrPheHisIleIleGluProIleTyr 501
      ::::::::::::::::::::
2793 GCAGAAATAATCCCTGAGCAGCAGCAGGATTGCACCAATTTAT 2837
      ::::::::::::::::::::

```

seq\_name: /sids5/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.ABA03041

seq\_documentation\_block:

ID ABA03041 standard; DNA; 2944528 BP.

AC ABA03041;

DT 05-FEB-2002 (first entry)

DE Listeria monocytogenes EGD-e genome sequence.

KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;

KW vitamin B12; bacterial infection; disease; ds.

OS Listeria monocytogenes.

PN W0200177335-A2.

PD 18-OCT-2001.

PF 11-APR-2001; 2001WO-FR01118.

PR 11-APR-2000; 2000FR-0004629.

PA (INSP) INST PASTEUR.

PI Buchrieser C, Frangeul L, Couve E, Rusnok C, Fsihi H, Deboux P;  
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;  
PI Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;  
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;  
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durand L;  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;  
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
PI Rose M, Voss H;

DR WPT; 2002-010914/01.

PT Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment  
PT and prevention of *Listeria* and related bacterial infections, and  
PT related polypeptides

PS Claim 1: SEQ ID No 1; 192pp; French.

CC The present sequence is the genome sequence of *Listeria monocytogenes*  
CC EGD-e. This sequence and fragments of this sequence are useful for  
CC selecting probes and primers for detecting genes in *L. monocytogenes* and  
CC related organisms, and to study genetic polymorphisms and other genomes.  
CC Proteins (AB847297-AB850149) expressed from the present sequence are  
CC useful for raising specific antibodies, identification of *L.*  
CC *monocytogenes* and related organisms, and for biosynthesis and  
CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and  
CC proteins encoded by it are also useful for selecting compounds that  
CC regulate gene expression and cell replication and modulate *L.*  
CC *monocytogenes*-related diseases. In addition, this sequence and proteins  
CC encoded by it are useful in pharmaceutical and vaccine compositions for  
CC the treatment or prevention of infections by *L. monocytogenes* and related  
CC organisms.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

```

alignment_scores:
  Quality: 470.50      Length: 579
  Ratio: 1.368        Gaps: 20
  Percent similarity: 59.413  Percent identity: 25.043
alignment_block:
US-09-824-567-2 x ABA03041/rev ..
Align seg 1/1 to reverse of: ABA03041 from: 1 to: 2944528
1 MetArgLysIleSerValGlyIleCysIleThrIleLeuLeuSerLeuSe 17
  ::::::::::::::::::::
2286215GTGAAAAATCTAATATTATTTCTTACACTTGGATTAACTTAACTTAA 2286166
17 rValValLeuGlnGlyCys.....LysG 25
  ::::::::::::::::::::
2286165CTTAGTCTTGGTAGCGGAGCGGATCCAGATCCAAATCCGACAAA 2286116
25 LysSerHisSerThrSerArgGlyGluLeuAlaIleAsnIleArg 41
  ::::::::::::::::::::
2286115AAGGCTCAGATTAGGAAAGCTTCAGGAGAGCAAGTAACTTAACTTACA 2286066
42 AspGluPro.....ArgSerLeuAspProArgGlnValArgLeuLeuSe 56
  ::::::::::::::::::::
2286065GAAAGCGCCTAATCTCTCTGCAGACAGCACAAAGCGGATGACCAAGT 2286016
56 rGluIleSerLeuValHisIleTyrGluGlyLeuValGlnGluAsnA 73
  ::::::::::::::::::::
2286015TGGTTTGAACGTTGTAACAAACAAAGAGGCTCTATATGCG..... 2285973
73 snLeuSerGlyAsnIleGluProAlaLeu.....AlaGluAspTyr 86
  ::::::::::::::::::::
2285972..CTTGACAAAGATGTTATCTCTGCCATTGCGCGTCTGCTGAAGAGCCA 2285925
87 SerLeuSerSerAspGlyLeuThrTyrThrPheLysLeuLys.....SerAl 102
  ::::::::::::::::::::
2285924AAATAGGAGTACACAAACAGTTTACTATCAACCTTCGGAAGATGC 2285875
102 aPheTyrSerAsnGlyAspProLeuThrAlaGluAspPheIleGluSerT 119
  ::::::::::::::::::::
2285874AAATGGTCAACCGAGGAGCTTAACTGCAATGACTATGTTACTCAT 2285825
119 rLysGlnValAlaThrGlnGluValSerGlyIleTyrAlaPheAlaLeu 135
  ::::::::::::::::::::
2285824GGCGTCTGCGCTTACCTTAACACTGCTGCAACATATTTACTTACTATT 2285775
136 AsnProIleLysAsnValArgLysIleGlnGluGlyHisLeuSerIleAs 152
  ::::::::::::::::::::
2285774GATGCAATCAAAAACGGTGGAGATATCTAGTGGCAAGAAAAACCTGA 2285725
152 pHisPheGlyValHisSerProAsnGluSerThrLeuValValThrLeuG 169
  ::::::::::::::::::::
2285724AGAAATAGGAATTAAGACGATAGATTATATTACTTTAGAACTTACTCTT 2285675
169 LysSerProThrSerHisPheLeuLysLeuAlaLeuProValPhePhe 185
  ::::::::::::::::::::
2285674CTAAACCACTGCTTACTTAACTCACTTTCGCAATTCCTCCCAACTTCTTC 2285625
186 ProValHis.....LysSerGlnArgThrLeuGlnSe 196
  ::::::::::::::::::::
2285624CCACTTAACGAAAAAATTCGTACGGAAGGCGGAAAAATATGACAAAA 2285575
196 rLysSerLeuProIleAlaSerGlyAlaPheTyrProLysAsnIle.... 211
  ::::::::::::::::::::
2285574TAGGTAAACATTTTCAATGGACCTTCGAGTTCAAGATTTGGACTG 2285525
212 ..LysGlnLysGlnTrpIleLysLeuSerLysAsnProHisTyrTyrAsn 227
  ::::::::::::::::::::
2285524GACACAAACAAAATGGACTTACGTA...AAAAATGATAAATATTGGAT 2285478
228 GlnSerGlnValGluThrLysThrIleThrIleHisPheIleProAspAl 244

```

```
2285477AAAGATAAAGTTAAGTTGAACAATAACAGCTACAGCTTCTTCACAGCTC 2285428
244 aAsnThrAlaAlaLysLeuPheAsnGlnGlyLysLeuAsnTrpGlnGlyP 261
2285427CGGTACTGGACCTTAACCTATATACAACT..... 2285400
261 ropTrpGlyGluArgIleProGlnGluThrLeuSer..... 273
2285399.....GACAAAGTTGACCGTACTGTATTGAGTCACACTATGCT 2285361
274 .....AsnLeuGlnSerLysGlyHisLeuHisSerPheAspValAlaG1 288
2285360GCTCAAAACAAACAAATAAAGACTATGTTACAGTGAATGAT...TCTTC 2285314
288 yThrSerTrpThrPheAsnIleAsnLysPheProLeuAsn..... 302
2285313TACTTCTATATAATTATACCAAAACGCTGCTGGCAAGACACAGTAT 2285264
303 .....AsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGlu 317
2285263TTGCTAACAAAACACATCGTAAGCTATAGCTCTTCGCGATTGACAAACAA 2285214
318 AlaLeuValSerThrIlePheLeuGlyArgAlaLysThrAlaAspHisLe 334
2285213TCGTATACTGATACAGTCTTAAACACGGATCTAAACCTGCAACAACT 2285164
334 uLeuProThr.....AsnIleHisSerTyrProG 344
2285163TGTACAGAGGTTTCACCTTTTCACCCAGCTAACAAAGAGATTATACAA 2285114
344 luHisGlnLysGlnGluMetAlaGlnArgGlnAlaTyrAlaLysLysLeu 360
2285113AGAAATCTGGCAAAACATTTAGAATATGATGTAAGAGAGCTCAAAAGCA 2285064
361 PheLysGluAlaLeuGluGluLeu.....GlnI1 370
2285063TGGAAAGCTGTTTGAAGAGATTAGAGTTAACGAAATTACAGTTGAATT 2285014
370 eThrAlaLysAspLeuGluHisLeuAsnLeuIlePheProValSerSerS 387
2285013CACTAGTGATGACTGAA.....AATGCCA 2284988
387 erAlaSerSerLeuValGlnLeuIleArgGluGlnTrpLysGluSer 403
2284987GAAATCTCT.....GAATTCATTCAAGACCACTACAAAAGAAC 2284947
404 Leu...GlyPheAlaIleProIleValGlyLysGluPheAlaLeuLeuG1 419
2284946TTAGATGCTTACAGTTAACTTAAACCTTAAACCTACCATTCAAGTTCGTTT 2284897
419 nAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyGlyTrpPheA 436
2284896CAAAATGACCAAAACCAAGATTACGATTCTCTATGACGCGCTGGGTC 2284847
436 laAspPheAlaAspProMetAlaPheLeuThrIlePheAlaTyrProSer 452
2284846CTGACTACTCAAGATCCATCTCTTCTTAGATCTATTTCGTAATGATGCT 2284797
453 GlyValProProTyrAlaIleAsnHisLysAspPheLeuGluIleLeuG1 469
2284796GCACAAAACAGATAGTATTCTTAACAAAGACTACGACAGATTTTAA 2284747
469 n.....AsnIleGluGlnGluAspHisGlnLysArg...SerGluL 483
2284746GTAGCTTCTGTAACCTTATGCGAGCTGACGATCAAAACCTGGGATGAA 2284697
483 euValSerGlnAlaSerLeuTyrLeuGluThrPheHisIleIleGluPro 499
2284696TGGTTTAAACAGCAAAAAATCTCTTACGGATGATGATGATGATGATGAT 2284647
500 IleTyrHisAspAlaPheGlnPheAlaMetAsnLysLysLeuSerAsnLe 516
```

```
2284646CTTTATCAAGCTTCTACTGATACCTACAAAAGACTACATTAAAAACTT 2284597
516 uGlyValSerProThrGlyValValAspPheArgTyr 528
2284596GCAAAAAAATCCATTGGT...CCAGATTACACTTTAC 2284563
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2002, 04:36:22 ; Search time 87.81 Seconds  
(without alignments)  
1048.096 Million cell updates/sec

Title: US-09-824-567-2  
Perfect score: 2739  
Sequence: 1 MRKISVGICIFILLSLSVL.....LSNLGVSPGVDFRYAKEN 532

Scoring table: BLOSUM62

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL19.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_virus.\*

16: sp\_bacteriaph.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2739	100.0	532	16 Q928Y9	Q928Y9 chlamydia p
2	1364	49.8	518	16 Q84201	Q84201 chlamydia t
3	1345	49.1	520	16 Q9PKJ4	Q9PKJ4 chlamydia m
4	1160.5	42.4	527	16 Q928Y8	Q928Y8 chlamydia p
5	755.5	27.6	526	16 Q928Y6	Q928Y6 chlamydia p
6	755.5	27.6	528	16 Q9K244	Q9K244 chlamydia p
7	748	27.3	527	16 Q9PKL9	Q9PKL9 chlamydia p
8	685.5	25.0	529	16 Q84178	Q84178 chlamydia m
9	536.5	19.6	550	16 Q97D46	Q97D46 clostridium
10	509.5	18.6	528	2 Q9K5F7	Q9K5F7 borrelia ga
11	506	18.5	528	2 Q9K5J9	Q9K5J9 borrelia af
12	505.5	18.5	528	2 Q31305	Q31305 borrelia bu
13	505.5	18.5	528	16 Q54584	Q54584 borrelia bu
14	482.5	17.6	435	16 Q928Y7	Q928Y7 chlamydia p
15	479.5	17.5	532	16 Q9AKR0	Q9AKR0 rhizobium m
16	478	17.5	559	16 Q929H6	Q929H6 listeria in

17	476	17.4	550	2 Q93QH8	Q93QH8 lactococcus
18	474.5	17.3	558	2 Q9LAT7	Q9LAT7 listeria mo
19	472	17.2	550	16 Q9CIL3	Q9CIL3 lactococcus
20	471.5	17.2	530	16 Q31315	Q31315 borrelia bu
21	467.5	17.1	514	2 Q53480	Q53480 borrelia co
22	467	17.1	545	16 Q9CIL2	Q9CIL2 lactococcus
23	446.5	16.3	553	16 Q927S4	Q927S4 listeria in
24	446	16.3	551	2 Q9FS06	Q9FS06 bacillus th
25	446	16.3	565	16 Q97D48	Q97D48 clostridium
26	440.5	16.1	529	2 Q9FLJ4	Q9FLJ4 enterococcus
27	440.5	16.1	519	2 Q9RMW8	Q9RMW8 bacillus an
28	431	15.7	543	16 Q9K114	Q9K114 vibrio chol
29	429.5	15.7	523	2 Q31313	Q31313 borrelia bu
30	429.5	15.7	529	2 Q31303	Q31303 borrelia bu
31	428.5	15.6	529	16 Q50927	Q50927 borrelia bu
32	428.5	15.6	530	16 Q985N6	Q985N6 rhizobium l
33	427.5	15.6	523	2 Q52615	Q52615 borrelia bu
34	425.5	15.5	529	16 Q51307	Q51307 borrelia bu
35	421	15.4	549	16 Q92E66	Q92E66 listeria in
36	419	15.3	426	16 Q84141	Q84141 chlamydia t
37	415.5	15.2	541	2 Q52616	Q52616 borrelia bu
38	414.5	15.1	541	16 Q51308	Q51308 borrelia bu
39	412	15.0	428	16 Q9PKP9	Q9PKP9 chlamydia m
40	412	15.0	549	16 Q9C127	Q9C127 lactococcus
41	411.5	15.0	523	2 Q31304	Q31304 borrelia bu
42	409.5	15.0	541	2 Q31306	Q31306 borrelia bu
43	400.5	14.6	547	16 Q83594	Q83594 treponema p
44	390.5	14.3	545	2 Q51643	Q51643 enterococu
45	379	13.8	522	2 Q06515	Q06515 streptococc

## ALIGNMENTS

RESULT 1

Q928Y9 ID Q928Y9 PRELIMINARY; PRT; 532 AA.  
AC Q928Y9;  
DT 01-MAY-1999 (TREMBL) 10, Created)  
DT 01-MAY-1999 (TREMBL) 10, Last sequence update)  
DT 01-DEC-2001 (TREMBL) 19, Last annotation update)  
DE OLIGOPEPTIDE BINDING PROTEIN.  
GN OPPA\_1 OR CPN0195 OR CP0572.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83358;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CWL029;  
RX MEDLINE=99206606; PubMed=10192388;  
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";  
RL Nat. Genet. 21:385-389(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AR39;  
RX MEDLINE=20150255; PubMed=10684935;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
pneumoniae AR39";  
RL Nucleic Acids Res. 28:1397-1406(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=J138;  
RX MEDLINE=20330349; PubMed=10871362;  
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138

RT from Japan and CWL029 from USA."  
 RL Nucleic Acids Res. 28:2311-2314 (2000).  
 DR EMBL: AE001606; AAD18348.1; -  
 DR EMBL: AE002216; AAF38391.1; -  
 DR EMBL: AF002545; BAA98405.1; -  
 DR TIGR: CF0572; -  
 DR InterPro: IPR000914; SBP\_bac\_5.  
 DR Pfam: PF00496; SBP\_bac\_5; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 532 AA; 59744 MW; 1CB473D9D46A1579 CRC64;

Query Match 100.0%; Score 2739; DB 16; Length 532;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-196;  
 Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MRKISVGICITILLSVLQCKSSHSSTSRGELAINIRDEPRSLDPRQVRLLS 60  
 DB 1 MRKISVGICITILLSVLQCKSSHSSTSRGELAINIRDEPRSLDPRQVRLLS 60  
 QY 61 VKHIYEGVQENNLGNIEPALAEDYSLSDGLTYTFKLKSAFWSNGDPLTAEDFI 120  
 DB 61 VKHIYEGVQENNLGNIEPALAEDYSLSDGLTYTFKLKSAFWSNGDPLTAEDFI 120  
 QY 121 QVATQVSGIYAFALNPINVKRIQEGHLSIDHFGVHSPNESTLVLTLESPTSHFLKLLA 180  
 DB 121 QVATQVSGIYAFALNPINVKRIQEGHLSIDHFGVHSPNESTLVLTLESPTSHFLKLLA 180  
 QY 181 LPVFPVHKQSRTLSQSLPIASGAFYPKNIKOKWIKLSKNPHYNNQSQVETKTIHF 240  
 DB 181 LPVFPVHKQSRTLSQSLPIASGAFYPKNIKOKWIKLSKNPHYNNQSQVETKTIHF 240  
 QY 241 IPDANTAAKLFNQGKLNQGWPGWGERIPQETLSNLSQKGLHSLFSDVAGTSLWTENKFP 300  
 DB 241 IPDANTAAKLFNQGKLNQGWPGWGERIPQETLSNLSQKGLHSLFSDVAGTSLWTENKFP 300  
 QY 301 LNNMKREALASALDKEALVSTIFLGRKTAADHLLPTNTHSHPEKQEMAQRAYAKKL 360  
 DB 301 LNNMKREALASALDKEALVSTIFLGRKTAADHLLPTNTHSHPEKQEMAQRAYAKKL 360  
 QY 361 FKEALELQITAKDLEHNLIPVSSASSLLVQLIREQWKESLGFAPIVGKFALLQA 420  
 DB 361 FKEALELQITAKDLEHNLIPVSSASSLLVQLIREQWKESLGFAPIVGKFALLQA 420  
 QY 421 DLSSGNFSLATGWFADPADMAFLTFAYPSGVPPYAINHKDFLEILQNTQEQQDQHR 480  
 DB 421 DLSSGNFSLATGWFADPADMAFLTFAYPSGVPPYAINHKDFLEILQNTQEQQDQHR 480  
 QY 481 SELVQSASLYLETFHIEPIYHDAFQFAMNKKLSNLGVSPGVVDPRYAKEN 532  
 DB 481 SELVQSASLYLETFHIEPIYHDAFQFAMNKKLSNLGVSPGVVDPRYAKEN 532

RESULT 2  
 O84201 PRELIMINARY; PRT; 518 AA.  
 ID O84201  
 AC O84201;  
 DT 01-NOV-1998 (TRENBLrel. 08, Created)  
 DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)  
 DE 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
 DE OLIGOPETIDE BINDING PROTEIN.  
 GN OPPA\_3 OR CT198.  
 OS Chlamydia trachomatis.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=813;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-D/UW-3/CX;  
 RX MEDLINE=99000809; PubMed=9784136;  
 RA Stephens R.S., Kalnan S., Lammei C.J., Fan J., Marathe R., Aravind L.,  
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
 RA Davis R.W.;  
 RT "Genome sequence of an obligate intracellular pathogen of humans:

RT Chlamydia trachomatis."  
 RL Science 282:754-759 (1998).  
 DR EMBL: AE001293; AAC67790.1; -  
 DR InterPro: IPR000914; SBP\_bac\_5.  
 DR Pfam: PF00496; SBP\_bac\_5; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 518 AA; 58858 MW; 8B3AE840831BBEF1 CRC64;  
 Query Match 49.8%; Score 1364; DB 16; Length 518;  
 Best Local Similarity 50.1%; Pred. No. 9.8e-94;  
 Matches 269; Conservative 97; Mismatches 147; Indels 24; Gaps 8;  
 QY 1 MRKISVGICITILLSVLQCKSSHSSTSRG-----ELAINIRDEPRSLDPRQVRLLS 56  
 DB 1 MRKISVGICITILLSVLQCKSSHSSTSRG-----ELAINIRDEPRSLDPRQVRLLS 56  
 QY 57 EISLVKHIEGLVQENNLGNIEPALAEDYSLSDGLTYTFKLKSAFWSNGDPLTAEDFI 116  
 DB 57 DINLIHLYEGVQETP-SGEVFPALAESEFLSEDKTYTFNLKKAFAWSNGDLITAHDFV 115  
 QY 117 ESKQVATQVSGIYAFALNPINVKRIQEGHLSIDHFGVHSPNESTLVLTLESPTSHFL 176  
 DB 116 RSNVDVLQNRVASIYSAFUPI-DVAK-----DSGFFAKDDHTLVINLLTPTPHFL 165  
 QY 177 KLALPVFPVH-KSQTLSQSLPIASGAFYPKNIKOKWIKLSKNPHYNNQSQVETK 235  
 DB 166 KLTLTPVFPVHSHQIRKEESLPISTGAFLEKRRWLEKLEKSPYYNNKQVAVQE 225  
 QY 236 ITHTIPDANTAAKLFNQGKLNQGWPGWGERIPQETLSNLSQKGLHSLFSDVAGTSLWTFN 295  
 DB 226 ICHIIIPQQAALFNQGLKQWGLPWGHSIPQETLATTNKRRAPRSPDISGTSNLFN 285  
 QY 296 INKFLNNMKREALASALDKEALVSTIFLGRKTAADHLLPTNTHSHPEKQEMAQRAYAKKL 355  
 DB 286 TAKKPFESHKLRQALSIVLNKEALASLAFV---APAKHLLPAHLHTYPPQSYKQOEAIT 342  
 QY 356 YAKKLFKALBELOITAKDLEHNLIPVSSASSLLVQLIREQWKESLGFAPIVGKEF 415  
 DB 343 LAKSLEELQITAKDLEHNLIPVSSASSLLVQLIREQWKESLGFAPIVGKEF 402  
 QY 416 ALQADLSSGNFSLATGWFADPADMAFLTFAYPSGVPPYAINHKDFLEILQNTQEQQ 475  
 DB 403 ALLQNDLIGNTFWMSGWGFADPADMAFLTFAYPSGVPPYAINHKDFLEILQNTQEQQ 475  
 QY 476 DHQKSELVQSASLYLETFHIEPIYHDAFQFAMNKKLSNLGVSPGVVDPRYAKEN 532  
 DB 462 NFQKRSALISEASLYIERQNVIEPLYHDFVHYTNNKLSFVLHPSGLVDMRYAKNS 518  
 RESULT 3  
 Q9PKJ4 PRELIMINARY; PRT; 520 AA.  
 ID Q9PKJ4  
 AC Q9PKJ4;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DE 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
 DE PEPTIDE ABC TRANSPORTER, PERIPLASMIC PEPTIDE-BINDING PROTEIN,  
 DE PUTATIVE.  
 GN TC0471.  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83560;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MOPN / NIGG;  
 RX MEDLINE=20150255; PubMed=1084935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwin M., Nelson W., DeBoy R., Kolony J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia



RT pneumoniae AB39. ;  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL; AE002315; AAF39319.1; -.  
 DR TIGR; TC0471; -.  
 DR InterPro; IPR000914; SBP\_bac\_5.  
 DR Pfam; PF00496; SBP\_bac\_5; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 520 AA; 59116 MW; 93E9C106CC0C7F58 CRC64;

Query Match 49.1%; Score 1345; DB 16; Length 520;  
 Best Local Similarity 49.4%; Pred. No. 2.6e-92;  
 Matches 266; Conservative 103; Mismatches 145; Indels 24; Gaps 9;

QY 1 MRKISVIGICITLLSVVLOCRESS-----HSSTSRGELAINIRDRPSLDPRQVRL 55  
 DB 1 MRKXVSGIC--LLIALAVITGCKSSKSNSSNSQ--SVSVKKDDPRFDFPREVRL 57  
 QY 56 SEISLVKHYVEGLVOENNSNIEPALAEDYSSDGLTYTFKLKSAFWSNGDPLTAEDF 115  
 DB 58 SDINLHLYVEGLVQETP-SGEVFPALAEFFLSEDKTYTFHLKALWSNGDLTAHDF 116  
 QY 116 IESKQVATQVSGIYAFALNPKVNRKIQEGHLSIDHFGVSNESLTVLTLESPTSIF 175  
 DB 117 VRSWDLQNRITASTYSAPFLPI-----DLSKO-CGFFAKDNHTLVINLHTPTPH 166  
 QY 176 LKLLALPVFPVH-KSQTLOSLSPLTASCAFYPNKIKOKWIKLSKNPHYNSOVETK 234  
 DB 167 LKULLPLVPVPHPEHQRNEAKALPSTGACLEKEDRWALKENPYNNKEOVAIQ 226  
 QY 235 TITIFIPDANTAAKLNCKGNQWPPGGERIPOETLSNLSQKGLHSHFDVAGTSWLT 294  
 DB 227 EIHIVDPQQTASALPKQKLDQGLPWSHISPOEALATANKRRTPQSFDSIGSWLT 286  
 QY 295 NINKPPLNMKRLASALDKALYSTIFIGRAKTADHLLPTNIHSYPEHOKQMAQRO 354  
 DB 287 NTSKVFFSHPKLQALSLVNLKALASTPV---KPAKHLLPLHLLTYPEQPTYKQEA 343  
 QY 355 AYAKLKEALEEIQITAKDLEHNLITPVSSASSLLVLIREQWKSIGLPAIPVQKE 414  
 DB 344 ILAKTLQEAULTDLNMTIKDLKCELFISATSSVNSQMAQMDQWRILGTFPTICKE 403  
 QY 415 FALQADLSSGNFSLATGWFADPADPMAFLTIFAYPSGVPVPAIHNKDFLEILONIEQ 474  
 DB 404 YALLQNDLTTTFEMSGWFAFSDPLAFUSVFS-SKIKPYALQDPLFLQILISITE 462  
 QY 475 QDHOKRSELYSQSLYETHTIIEPIYHDAFQFAMNKKLSNLGVSPTGVDFRYAKEN 532  
 DB 463 KDPKRSLISEASLYLEKQNIIEPLYHDVHYAANNKLSFVRLHPSGLVDMRYAKNS 520

RESULT 4  
 Q928Y8 PRELIMINARY; PRT; 527 AA.  
 AC Q928Y8  
 DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE OLIGOPETIDE BINDING PROTEIN.  
 GN OPA\_2 OR CPN0196 OR CP0571.  
 OS Chlamydia pneumoniae (Chlamydophila pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CW1029;  
 RX MEDLINE=99206606; PubMed=10192388;  
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
 RL Nat. Genet. 21:385-389(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Winn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RX MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CW1029 from USA.";  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 DR EMBL; AE001606; AAD18349.1; -.  
 DR EMBL; AE002215; AAF38390.1; -.  
 DR EMBL; AP002545; BAA98406.1; -.  
 DR TIGR; CP0571; -.  
 DR InterPro; IPR000914; SBP\_bac\_5.  
 DR Pfam; PF00496; SBP\_bac\_5; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 527 AA; 61166 MW; 628F32FB4F13D79E CRC64;

Query Match 42.4%; Score 1160.5; DB 16; Length 527;  
 Best Local Similarity 43.2%; Pred. No. 1.7e-78;  
 Matches 225; Conservative 111; Mismatches 176; Indels 9; Gaps 4;

QY 15 SLSVLOCCKESSHSTSRGELAINIRDRPSLDPRQVRLSEISLVKHYVEGLVQENNL 74  
 DB 11 TLMWLTSGC---SPSQSGKGFVVMKEMPRSLDPCKTFLADQTLMRHLYVEGLVEHQ 67  
 QY 75 SGNIEPALAEDYSSDGLTYTFKLKSAFWSNGDPLTAEDFIESKQVATQVSGIYAF 134  
 DB 68 NGIKPALAESTTISEGTRYTFKIKNLSNGDPLTAQDFVSSWKEILKEDASSVLYA 127  
 QY 135 LNPKNVRKIQEGHLSIDHFGVSNESLTVLTLESPTSHELKLLALPVFPVHKSORTL 194  
 DB 128 FLPIKNARAIFDDTESPENLGVRLDKRHLQLETPCAHELFELTLPFPVPHETLRN 187  
 QY 195 QS--KSLPIASCAFYPNKIKOKWIKLSKNPHYNSQVETKTTTIFIPDANTAAKLFN 252  
 DB 188 STSFEMPTTCGAFRPSVLEKGLRHLKKNPMYHNKSVKLKTIQVTSNANTAILFK 247  
 QY 253 QGKLNQWGPMPGERIPOETLSNLSQKGLHSHFDVAGTSWLTFFNKPPLNNMKREALAS 312  
 DB 248 HKKLDQWGPWGEPIPETASLHDDQLFSLPGASTTWLLFNQKPNNAKRLKSL 307  
 QY 313 ALDKALYSTIFIGRAKTADHLLPTNIH--SYPEHOKQMAQVAKKLPKALEELQI 370  
 DB 308 AIDKMLTKVYQGLAEPDHLHRLPYGTYPERKQN--ERILEAQQLFEALDELQ 365  
 QY 371 TAKDLEHNLITPVSSASSLLVLIREQWKSIGLPAIPVGVKFEALQADLSSGNFSLA 430  
 DB 366 TREDEKETLTFSTFSYGRICQMLREQWKKVLKTIPIVQGEFTTQKNFLBNLSLT 425  
 QY 431 TGGWFADPADPMAFLTIFAYPSGVPVPAIHNKDFLEILONIEQDQHKRSELYSQSLY 490  
 DB 426 VNQWTAFTDPWSILMIFANPFGSIPYHLODSHFQTLKTIQEHKHLRLIIEALDY 485  
 QY 491 LEFTHIEPIYHDAFQFAMNKKLSNLGVSPTGVDFRYAKE 531  
 DB 486 LERCHILEPLCHPNLRNALNKNFNLVFRRTSDFRFLEK 526

RESULT 5  
 Q928Y6 PRELIMINARY; PRT; 526 AA.  
 ID Q928Y6

AC	Q928V6;
AD	01-MAY-1999 (TrEMBLrel. 10, Created)
AE	01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
AF	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
AG	OLIGOPEPTIDE BINDING PROTEIN.
AH	OPPA_4 OR CPN0198.
AI	Chlamydia pneumoniae (Chlamydia pneumoniae).
AJ	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
AK	NCBI_TaxID-83556;
AL	[1] SEQUENCE FROM N.A.
AM	RP STRAIN=CWL029;
AN	RC MEDLINE=99206606; PubMed=10192388;
AO	RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
AP	olinger L., Grimwood J., Davis R.W., Stephens R.S.;
AQ	RT "Comparative Genomes of Chlamydia pneumoniae and C. trachomatis."
AR	RL Nat. Genet. 21:385-389(1999).
AS	[2]
AT	RP SEQUENCE FROM N.A.
AV	RC STRAIN=J138;
AW	RX MEDLINE=20330349; PubMed=10871362;
AX	RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
AY	Shiba T., Ishii K., Hattori M., Kuhrata S., Nakazawa T.;
AZ	RA Comparison of whole genome sequences of Chlamydia pneumoniae J138
BA	from Japan and CWL029 from USA.;
BB	RT Nucleic Acids Res. 28:2311-2314(2000).
BC	RL EMBL: AE001606; RAD18351.1. -
BD	DR EMBL: AP002545; BAA98408.1. -
BE	DR InterPro: IPR000914; Srp_bac_5.
BF	DR Pfam: PF00496; Srp_bac_5; 1.
BG	KW Complete proteome.
BH	SEQUENCE 526 AA; 59822 MW; D149373CC92C18C CRC64;

Query Match	27.6%;	Score 755.5;	DB 16;	Length 526;
Best Local Similarity	33.9%;	Pred. No. 3.2e+48;		
Matches 177;	Conservative 106;	Mismatches 216;	Indels 23;	Gaps
QY 13	LLSLSVLQCKESSHSSTRGELAINIRDEPRSLDPROVRLSEISLVKHVYGLVQEN 72			
Db	15 LLFLTLTSCSCSQKQEPGLG-HLVAMSHDLADDPNAYLSRDSALAKAYIELGTRT 73			
QY 73	NLSGNIFPALAEDYSLSSDGLTYTFKLKAFWSNGDPLTAEDFIESKOVATQVSGIYA 132			
Db	74 D--OGIALALAEASYLTSDKHVYTFKLRPSVMSWDGTPLTAYDEFSIKQLTFEFPSPSIH 131			
QY 133	FALNPINKVRKTOEGLHSIDIEFGVHSPNESTLVLTSPSTHKLALPVPFVPHKSOR 192			
Db	132 TLGVYKNSAHHNAQKSLETGLQAKDDLTVLTQPPPYFLTLARPVSFVHTLR 191			
QY 193	TLOSLSLP----IASGAPYKNIQKQWIKLSKKNPHYNYOSQVETKTTITHFIDPANTA 248			
Db	192 ESYKKGTPPSTYISNGPVLKKEHONYLTLEKPHYVHDESHKLDRTVLKLPDASTAT 251			
QY 249	KLFNQCKLWNOGPPWGERIPQETLSNLOSKGHLASFQVAGTSWLTFTNFKNPPNNMKLRE 308			
Db	252 KLFKSIDWIGSPWAPISNED-QKVLISQEKILTYVSTLLIYNLQKPLIQNKALRK 310			
QY 309	ALASALDKALYSIFILGRKATADHLPLTNHSPYHQEQMAORQAYAKKLFKEALEEL 368			
Db	311 AIAHADRKSILRLVPSG--QEAVTLVPPNLSQLNQLQKELSTEERQYKARAYQAEAKTL 368			
QY 369	QITAKDLEHLNLIPVSSSASSLLVOLIREQWESLGFAIPWKEFALLQADLSSGNFS 428			
Db	369 --SEKELAEGLTLPIDSSNSTIIAQEIQRLQKDTGLAKIQGMEYHCFKKRRQGOFF 426			
QY 429	LATGGWFADFPADPAELTIFAYPSGVPYAINKDFLEILQNI---EQEDHOKRSSELVS 485			
Db	427 IATGWIAYEYSPVAFSLISGLNPRDLTQW--RNSDYETLEKVLPHAYKENLKAEMI- 483			
QY 486	QASLYLTFTHIEPIYHDAFQFAMNKLNLNLSVSGTVGVDFR 527			
Db	484 ----IEETIPLPHGYIYAIHPKIQNTFSLIGHTDLK 520			

```

RESULT      6
ID          Q9K244
AC          PRELIMINARY; PRT; 528 AA.
DT          01-OCT-2000 (TrEMBLrel. 15, Created)
DT          01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT          01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE          PEPTIDE ABC TRANSPORTER, PERIPLASMIC PEPTIDE-BINDING PROTEIN.
GN          CP0569.
OS          Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC          Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX          NCBI_TaxID=83558;
EN          [1] SEQUENCE FROM N.A.
RC          STRAIN=AR39;
RX          MEDLINE=20150255; PubMed=10684935;
RA          Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA          White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA          Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA          Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.
RA          Eisen J., Fraser C.M.;
RA          "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT          pneumoniae AR39."
RL          Nucleic Acids Res. 28:1397-1406(2000).
RD          EMBL: AE002215; AAF38388.1; -.
DR          TIGR: CP0569;
DR          InterPro: IPR000914; SBP_bac_5.
DR          Pfam: PF00496; SBP_bac_5; 1.
SQ          SEQUENCE 528 AA; 60082 MW; F2722GCCQCFDRI1 CP0564.

```

Query Match	27.68;	Score 755.5;	DB 16;	Length 528;
Best Local Similarity	33.99;	Pred. No. 3 2e-48;		
Matches 177;	Conservative 106;	Mismatches 216;	Indels 23;	Gaps
QY	13	LLLSVVLGGCKESHSTSGELAINIRDEPSRLDPQVRLLSEISLKHVIEGLVQEN 72		
DB		: : : : :       : : : : :       : : : : :       : : : : :		
QY	17	LLFLLTLSCSKQKQEPGLK-HLVITAMSHDLADLPNAVLSRDSALAKALYEGLPRET 75		
DB		: : : : :       : : : : :       : : : : :       : : : : :		
QY	73	NLSNTEPALAEDYSLSSDGLVTFKLLSAFNSGCDPLTAEDFIESWQVATQEVSGIYA 132		
DB		: : : : :       : : : : :       : : : : :       : : : : :		
QY	76	D-QGIALAESAIVLSKDHVYFKLRPSWSDGTLTADYFEKSIQLYFEFSPSTIH 133		
DB		: : : : :       : : : : :       : : : : :       : : : : :		
QY	133	FALNPKNVKYIQEHLISIDHGVHSPNESTLVLTSPSTSHFKLLALPVFPVHKSOR 192		
DB		: : : : :       : : : : :       : : : : :       : : : : :		
QY	134	TLAGVKNSSALHNAQKSLTGLQAKDDLTVITLEQFPFLLIARPVSPVHHTLR 193		
DB		: : : : :       : : : : :       : : : : :       : : : : :		
QY	193	TLQKSLP----IASGAFYKNTKQWIKLSKNPHYNYNOSQVETKTIHIFPDANTAA 248		
DB		: : : : :       : : : : :       : : : : :       : : : : :		
QY	194	ESYKKTPTPSYISNGPVLKHEQNTLILEKNPHYNYDHESVKLDRTLKLIIPDASTAT 253		
DB		: : : : :       : : : : :       : : : : :       : : : : :		
QY	249	KLFNOGKNWQGPWGERIPQETLSNLOSKGLHSPDVAGTSWLTFFNKNFPLNNKKLE 308		
DB		: : : : :       : : : : :       : : : : :       : : : : :		
QY	254	KLFKSKIDWIGSPWSPAPISNEQKVLQSKELITYSVSTLLIYNLOKPLQNKALR 312		
DB		: : : : :       : : : : :       : : : : :       : : : : :		
QY	309	ALASALDEALVSTFIFIGRAKTAADHLPTNHSYPEHQEQMAORQAVAKLKEALEL 368		
DB		: : : : :       : : : : :       : : : : :       : : : : :		
QY	313	ATAHAIDRKSLRLVPSG--QBAVTLVPPNLSQLNLQKEISTEERQTKARAYQEAKEVTL 370		
DB		: : : : :       : : : : :       : : : : :       : : : : :		
QY	369	QITADLLEHMLNIPVSSASSLLVOLIROWKESIGFAIPVIGKEFALLQADLSGNSF 428		
DB		: : : : :       : : : : :       : : : : :       : : : : :		
QY	371	--SEKEIAEUSILYPIDSSNSIIAQEIQRLQMDTLGLKIKIQGMEVHCFLKRRQGDFF 428		
DB		: : : : :       : : : : :       : : : : :       : : : : :		
QY	429	LATGWQFADEADPWAFLTIFAPVSGVPVYAINKHOFLEILONI---EQDQHQKRSSELVS 485		
DB		: : : : :       : : : : :       : : : : :       : : : : :		
QY	429	IATGQIAEIVSPVRLSILNGPDLQW--RNSDYETKLEKLYPHAYKENLKAEMI- 485		
DB		: : : : :       : : : : :       : : : : :       : : : : :		
QY	486	QASLYLTFTHIEPYDHAOFAMNKKLSNLGVSPTGVVDPER 527		
DB		: : : : :       : : : : :       : : : : :       : : : : :		
QY	486	-----IEEPTPIPLVHGKYIYATPKIOMTNGSLGHTYK 529		
DB		: : : : :       : : : : :       : : : : :       : : : : :		

## RESULT 7

Q9PKL9  
ID Q9PKL9 PRELIMINARY; PRT; 527 AA.  
AC Q9PKL9  
DT 01-OCT-2000 (TREMELrel. 15, Created)  
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)  
DE PEPTIDE ABC TRANSPORTER, PERIPLASMIC PEPTIDE-BINDING PROTEIN.  
GN TC0446.  
OS Chlamydia muridarum.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83560;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MOPN / NIGG;  
RX MEDLINE=20150255; PubMed=10684935;  
RA White D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA Linher K., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
RA Eisen J., Fraser C.M.;  
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
pneumoniae AK39.";  
RL Nucleic Acids Res. 28:1397-1406(2000).  
DR EMBL: AE002313; AAF39300.1; -  
DR TIGR: TC0446; -  
DR InterPro: IPR000914; SBP\_bac\_5.  
DR Pfam: PF00496; SBP\_bac\_5; 2.  
KW Complete proteome.  
SQ SEQUENCE 527 AA; 60452 MW; 510E8C6B49BF437 CRC64;

Query Match 27.3%; Score 748; DB 16; Length 527;  
Best Local Similarity 35.3%; Pred. No. 1.2e-47;  
Matches 185; Conservative 96; Mismatches 217; Indels 26; Gaps 12;  
QY 14 LSLSVVLOGKESHSSTSRGEIAINIRDEPRSLDPRVRLSEISLVKHIVGVLQENN 73  
DB 14 LSLTGLTSCHOKEEN--LRNIRVAICHDPMSLDPRQVFLKDVICAKALYEGVLREND 71  
QY 74 LSGNTEPALAEDYSLSDGLTYTFKLKSAFWSNGDPLTAEDFTESMKQVATQVSGIYAF 133  
DB 72 --GSHLALAEYSLSDRCVYTFKLTFWHNGDLVATYDFEESIKQYLHEVDNVALR 129  
QY 134 ALNPKIKVRKIQEGHLSIDHFGVHSPNESTLVVLTESPTSHFLKLLALPVFPVHKSS--- 190  
DB 130 LLALIKNSHAVLGDIPVENLGVRLDDEHLEITLHPSSHSELETLPVFPVHSLRE 189  
QY 191 -QRTLOSLSPIAS--GAFYPKNIKOKWIKLSKNPHYNQSOVETKTITIHFPDANTA 248  
DB 190 YRNRSKRLPIISNGPFIIRCYEPQNFLLDKNPFYHQKNSLDAYRLQIVPDIHTAV 249  
QY 249 KLFNOGKLNWQGPWGERIPQETLSNLSQSGHLSHSDVAGTSWLTFTNINKPPLNNMKLR 308  
DB 250 QLFCKKYVDVLGLPSSSPFLBEQKNL--SQDFLYDYPVNLCTVLCFNVNHPKLPDPSLRA 308  
QY 309 ALASALDKALYSTIFLGRKTAADHLLPTNHSYPEHQEQMAQQAQYAKLFEALEEL 368  
DB 309 ALSLAIDRETLK--LACKGSIATSFVHPSLSKMLDVLQS--KERISLANNALAEALK-- 363  
QY 369 QITAKOLEHNLIFPVSSASSLLVLIREQWKESLGFAIPVIGKREFALLQADLSGNSFS 428  
DB 364 TVPQEELEKITLTIPIESIVLRAVQVIRQQLFDVLGFKISTLGLVHSPFLDKRSRGEFS 423  
QY 429 LATGGWFADFPMAFLTIFAYSGVPPYAINHKDFLET-----LQNIQEQQDHOKRS 481  
DB 424 LSTGNWVADYQQAFLSLGNGTKYQDVIDWQNOEYTDIVSRLLVEDSTDLQILAE- 482  
QY 484 VSOASLYLETHIEPIYHDAFQFAMNKKLSNLGVSPTGVVDFR 527  
DB 483 ----QILLKESPLI-PLYHLDYAYAKHPKVSNLQTSLSIGEIDLK 521

## RESULT 8

O84178  
ID O84178 PRELIMINARY; PRT; 529 AA.  
AC O84178;  
DT 01-NOV-1998 (TREMELrel. 08, Created)  
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)  
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)  
DE OLIGOPEPTIDE BINDING PROTEIN PERWEASE.  
GN OPA\_2 OR Cti75.  
OS Chlamydia trachomatis  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=813;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=D/UW-3/CX;  
RX MEDLINE=99000809; PubMed=9784136;  
RA Stephens R.S., Kallam S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
RA Davis R.W.;  
RT "Genome sequence of an obligate intracellular pathogen of humans:  
Chlamydia trachomatis.";  
RL Science 282:754-759(1998).  
DR EMBL: AE001291; AAC67766.1; -  
DR InterPro: IPR000914; SBP\_bac\_5.  
DR Pfam: PF00496; SBP\_bac\_5; 2.  
KW Complete proteome.  
SQ SEQUENCE 529 AA; 60331 MW; 2D546BD1DFC3F786 CRC64;

Query Match 25.0%; Score 685.5; DB 16; Length 529;  
Best Local Similarity 34.4%; Pred. No. 5.5e-43;  
Matches 181; Conservative 89; Mismatches 227; Indels 29; Gaps 10;  
QY 14 LSLSVVLOGKESHSSTSRGEIAINIRDEPRSLDPRVRLSEISLVKHIVGVLQENN 73  
DB 14 VSLAFLGTLSCYHKKEE--PKDVLRTAICHDPMSLDPRQVFLSKDVSIVKALYEGVLRE 71  
QY 74 LSGNTEPALAEDYSLSDGLTYTFKLKSAFWSNGDPLTAEDFTESMKQVATQVSGIYAF 133  
DB 72 AA--FQLALAEYHQSDGCVYTFKLTFWHNGDLVATYDFEESIKQYFREIDNPSLR 129  
QY 134 ALNPKIKVRKIQEGHLSIDHFGVHSPNESTLVVLTESPTSHFLKLLALPVFPVHKSSORT 193  
DB 130 SLALIKNSHAVITGALPVEDIGVRLNAKTEILEIVLENFPYFLEILAHVPVFPVHTSLRE 189  
QY 194 L-----QSKSLPIAS--GAFYPKNIKOKWIKLSKNPHYNQSOVETKTITIHFPDANTA 247  
DB 190 YKDKRNRKRPFIISNGPFAIQCYEPQVRYLLINKNPLYHAKHVDLNSVCLQIVPDIHTA 249  
QY 248 AKLFNOGKLNWQGPWGERIPQETLSNLSQSGHLSHSDVAGTSWLTFTNINKPPLNNMKLR 307  
DB 250 MGLFQKNHIDVLGLPSSSPFLBEQKNL--PREKLFDPVLSVLCFNIHQTPLNPSLR 308  
QY 308 EALASALDKALYSTIFLGRKTAADHLLPTNHSYPEHQEQMAQQAQYAKLFEALEE 367  
DB 309 TALSIAINRETLK--LACKGSIATSFVHPSLSKMLDVLQS--DERIALAKGYLEALKT 365  
QY 368 LQITAKOLEHNLIFPVSSASSLLVLIREQWKESLGFAIPVIGKREFALLQADLSGNSNF 427  
DB 366 L--SQEDELKITLTIPIESVCLRAVQVIRQQLFDVLGFKISTLGLVHSPFLDKRSRGEF 423  
QY 428 SLATGGWFADFPMAFLTIFAYSGVPPYAINHKDFLET-----LQNIQEQQDHOKRS 481  
DB 424 SLATGNWVADYHQASAFSLVGLG-----NGTRYKDFQLINQWQNKYINIVALLQIIESS 476  
QY 482 ELVSOASLYLETHIEPIYHDAFQFAMNKKLSNLGVSPTGVVDFR 527  
DB 477 DLQLMAEQLLLESPLIPLYHLDYAYAKHPKVSNDLQTSLSRGEIDLK 522

RESULT 9  
Q97D46

[illegible]

Db	520	YTNWVEKPYVKDLHKSPGLGFVYFNNTYIKK	550
RESULT	10		
Q9K5F7			
ID	Q9K5F7	PRELIMINARY;	PRT; 528 AA.
AC	Q9K5F7		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	SURFACE ANTIGEN.		
OS	P58.		
OS	Borrelia garinii.		
OC	Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.		
OX	NCBI_TaxID=29519;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=PBI;		
RA	Wilske B., Habermann C., Fingerle V., Hillenbrand B.,		
RA	Jauris-Helpe S., Lehnert G., Pradel I., Roessler D.,		
RA	Schulte-Spechtel U.;		
RT	"An improved recombinant IgG immunoblot for serodiagnosis of Lyme		
RT	borreliosis.";		
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AJ250013; CAB6045.1; -		
DR	HSP; P06202; IBS2		
DR	InterPro: IPR000914; SBP_bac_5.		
DR	Pfam: PF00496; SBP_bac_5; 1.		
SEQUENCE	528 AA; 60562 MW; AA6F2B802CBC46FR CPG64.		

```

Q9K5J9
ID Q9K5J9 PRELIMINARY; PRT; 528 AA.
AC Q9K5J9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SURFACE ANTIGEN.
GN P58.
OS Borrelia afzelii.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=29518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PKO;
RA Wilske B., Habermann C., Fingerle V., Hillenbrand B.,
RA Jauris-Helpe S., Lehnert G., Pradel I., Roessler D.,
RA Schulte-Spechtel U.;
RT "An improved recombinant IgG immunoblot for serodiagnosis of Lyme
RT borreliosis.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ250012; CAB6044.1; -.
DR HSSP; P06202; IBS2.
DR InterPro; IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5; 1.
SQ SEQUENCE 528 AA; 60509 MW; E2D048F7290A7A16 CRC64;

Query Match 18.5%; Score 506; DB 2; Length 528;
Best Local Similarity 28.0%; Pred. No. 1.4e-29;
Matches 139; Conservative 106; Mismatches 210; Indels 42; Gaps 14;

QY 24 KESSHSSTSRGELAINRDRSPQVRLSEISLVKHYEGLVQNNLSNIEPALA 83
DQ 24 KESSHSSTSRGELAINRDRSPQVRLSEISLVKHYEGLVQNNLSNIEPALA 83
DB 24 KESSHSSTSRGELAINRDRSPQVRLSEISLVKHYEGLVQNNLSNIEPALA 83
QY 21 KEKKEKGVSTFRLSEISLVKHYEGLVQNNLSNIEPALA 77
DQ 21 KEKKEKGVSTFRLSEISLVKHYEGLVQNNLSNIEPALA 77
DB 21 KEKKEKGVSTFRLSEISLVKHYEGLVQNNLSNIEPALA 77
QY 84 EDYSSDGLTYFLK-SAFWSNGDPLTAEDFIESWKQVATQVSGIYA-FALNPIKNV 141
DQ 84 EDYSSDGLTYFLK-SAFWSNGDPLTAEDFIESWKQVATQVSGIYA-FALNPIKNV 141
DB 84 EDYSSDGLTYFLK-SAFWSNGDPLTAEDFIESWKQVATQVSGIYA-FALNPIKNV 141
QY 78 KSWIDSPDQVTVTLREKIIWSDGVAITAEGIRKSYLRILNKETGSGNYSVAVKSTIKNG 137
DQ 78 KSWIDSPDQVTVTLREKIIWSDGVAITAEGIRKSYLRILNKETGSGNYSVAVKSTIKNG 137
DB 78 KSWIDSPDQVTVTLREKIIWSDGVAITAEGIRKSYLRILNKETGSGNYSVAVKSTIKNG 137
QY 142 RKIQEGHLSIDHFGVHSPNESTLVVLESPTSHFLKLALPVP-PPVHKQSRTLSQSKL 199
DQ 142 RKIQEGHLSIDHFGVHSPNESTLVVLESPTSHFLKLALPVP-PPVHKQSRTLSQSKL 199
DB 142 RKIQEGHLSIDHFGVHSPNESTLVVLESPTSHFLKLALPVP-PPVHKQSRTLSQSKL 199
QY 138 QKVFQGVSDSELGIRADEKTEITLESPPKPYFIDMLVHQSFIPV-PIVHTKYGQNTSPENMVTSGPFKL 256
DQ 138 QKVFQGVSDSELGIRADEKTEITLESPPKPYFIDMLVHQSFIPV-PIVHTKYGQNTSPENMVTSGPFKL 256
DB 138 QKVFQGVSDSELGIRADEKTEITLESPPKPYFIDMLVHQSFIPV-PIVHTKYGQNTSPENMVTSGPFKL 256
QY 200 P---IASGAFYKNIKQKIKSLNHYNQSVQVETKTIITHTPDANTAAKLNQKMGPPWGERIP 197
DQ 200 P---IASGAFYKNIKQKIKSLNHYNQSVQVETKTIITHTPDANTAAKLNQKMGPPWGERIP 197
DB 200 P---IASGAFYKNIKQKIKSLNHYNQSVQVETKTIITHTPDANTAAKLNQKMGPPWGERIP 197
QY 198 PENIVTSGPFKLKERNEXYVVEKNDYNSNQVEQVEITFTNDNSYAYKMYENGEL 257
DQ 198 PENIVTSGPFKLKERNEXYVVEKNDYNSNQVEQVEITFTNDNSYAYKMYENGEL 257
DB 198 PENIVTSGPFKLKERNEXYVVEKNDYNSNQVEQVEITFTNDNSYAYKMYENGEL 257
QY 257 NQQGPWGERIPQETLSNLSQKHLHSFVAGTSMLTNFKNPKLNKMLREALASALDK 316
DQ 257 NQQGPWGERIPQETLSNLSQKHLHSFVAGTSMLTNFKNPKLNKMLREALASALDK 316
DB 257 NQQGPWGERIPQETLSNLSQKHLHSFVAGTSMLTNFKNPKLNKMLREALASALDK 316
QY 258 D----AIPSAIPDLKRLSDYSSAVNAIFYAFNTHIKPLDNVAVKRALTLAIDR 313
DQ 258 D----AIPSAIPDLKRLSDYSSAVNAIFYAFNTHIKPLDNVAVKRALTLAIDR 313
DB 258 D----AIPSAIPDLKRLSDYSSAVNAIFYAFNTHIKPLDNVAVKRALTLAIDR 313
QY 317 EALVSTIFLGRKTAADHLLPTNIHSGYPEHQKQMAQRAYAK--KLFKEALEELQITAKD 374
DQ 317 EALVSTIFLGRKTAADHLLPTNIHSGYPEHQKQMAQRAYAK--KLFKEALEELQITAKD 374
DB 317 EALVSTIFLGRKTAADHLLPTNIHSGYPEHQKQMAQRAYAK--KLFKEALEELQITAKD 374
QY 314 EFLTYKV-LDGTTPTRTPNPFSSY-----SYAKNLEFNPETAKTLAEG 360
DQ 314 EFLTYKV-LDGTTPTRTPNPFSSY-----SYAKNLEFNPETAKTLAEG 360
DB 314 EFLTYKV-LDGTTPTRTPNPFSSY-----SYAKNLEFNPETAKTLAEG 360
QY 375 LEHNLIPV-----SSASALLVQLIREQWESLGFAIPVKGFEALLQADLSSGNFS 428
DQ 375 LEHNLIPV-----SSASALLVQLIREQWESLGFAIPVKGFEALLQADLSSGNFS 428
DB 375 LEHNLIPV-----SSASALLVQLIREQWESLGFAIPVKGFEALLQADLSSGNFS 428
QY 361 YPNGN-GPPIKLYNTSEAHKICEFQNGWKILNIDVELENEEWITLYLNKNSNGYE 419
DQ 361 YPNGN-GPPIKLYNTSEAHKICEFQNGWKILNIDVELENEEWITLYLNKNSNGYE 419
DB 361 YPNGN-GPPIKLYNTSEAHKICEFQNGWKILNIDVELENEEWITLYLNKNSNGYE 419
QY 429 LATGWFADPADPMAFLTFAYPSGVPPYAINH---KDFLEILQIQEQDHQKRELSVS 485
DQ 429 LATGWFADPADPMAFLTFAYPSGVPPYAINH---KDFLEILQIQEQDHQKRELSVS 485
DB 429 LATGWFADPADPMAFLTFAYPSGVPPYAINH---KDFLEILQIQEQDHQKRELSVS 485
QY 420 IARAGWIGDYADPLTFLSILTH--GYTFSSSHNSYSPYENELIKKSDLELDPKIRQDILR 477
DQ 420 IARAGWIGDYADPLTFLSILTH--GYTFSSSHNSYSPYENELIKKSDLELDPKIRQDILR 477
DB 420 IARAGWIGDYADPLTFLSILTH--GYTFSSSHNSYSPYENELIKKSDLELDPKIRQDILR 477
QY 486 QA-SLYLETFHTEPIY 501
DQ 486 QA-SLYLETFHTEPIY 501
DB 486 QA-SLYLETFHTEPIY 501
QY 478 KAEIIEKDFIPIY 494
DQ 478 KAEIIEKDFIPIY 494
DB 478 KAEIIEKDFIPIY 494

RESULT 12
ID O31305 PRELIMINARY; PRT; 528 AA.
AC O31305;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)

```





Db 216 AIIHLNRGKVDWQFVHOGIPWEL--HKOSQYHYTYVEGAFWLCINTKSPHLNDLQN 273  
 QY 307 REALASALDEALVSTIFLGRKATADHLLPTNIHSYPERHOKQEMADQAYAKKLFKEALE 366  
 Db 274 RRLATCIDKRSIIEEALGGTQOAEFL-----SRGAPQPNQYKK----- 313  
 QY 367 ELQITAKDLHLNLIFFPVSSASSLLVQLIREQWKESLGFAPIVGKEFALLQADLSSGN 426  
 Db 314 --QKPLTPQEKLVIVPSDILRCORIAEILKEQWK-AAQIDILLEGLEYHLFVNRKRVQD 370  
 QY 427 FSLATGGWADPADMAELTITFVPSGVPYAINHKDFLEILQNEQDQHKRSELVSQ 486  
 Db 371 YAIAT-----QTGVAYPG-----ANLISE 390  
 QY 487 ASLYLETFHIIPIYHDAFAMNKLNLGYSPTGVDFRY 528  
 Db 391 EDKLLQNFEII--PIYLSYDYLTDQFIEGVYINASGAVDLKY 431

RESULT 15  
 Q9AKR0  
 ID Q9AKR0 PRELIMINARY; PRT; 532 AA.  
 AC Q9AKR0;  
 DT 01-JUN-2001 (Tremblrel. 17, Created)  
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE OLIGOPEPTIDE ABC TRANSPORTER (PUTATIVE OLIGOPEPTIDE UPTAKE ABC  
 DE TRANSPORTER PERIPLASMIC SOLUTE-BINDING PROTEIN).  
 GN OPPA.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 CX NCBI\_TaxID=382;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=CR4;  
 RA Nogaes J.; Olivares J.; Sanjuan J.;  
 RT "Oligopeptide transport systems in Sinorhizobium meliloti";  
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=1021; FLASMID=PSYMB (MEGAPLASMID 2);  
 RX MEDLINE=21396508; PubMed=11481431;  
 RA Finan T.M.; Weidner S.; Wong K.; Buhmester J.; Chain P.;  
 RA Vorhoeiter F.J.; Hernandez-Lucas I.; Becker A.; Gouzy J.;  
 RA Golding B.; Puehler A.;  
 RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-  
 RT fixing endosymbiont Sinorhizobium meliloti";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).  
 DR EMBL; AJ296269; CAC33589.1;  
 DR EMBL; AL603645; CAC49261.1;  
 DR HSSP; P06202; 1B52.  
 DR InterPro; IPR000914; SBP\_bac\_5.  
 DR Pfam; PF00496; SBP\_bac\_5; 1.  
 KW Plasmid; Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 532 AA; 59004 MW; FA0A382E4027D691 CRC64;

Query Match 17.5%; Score 479.5; DB 16; Length 532;  
 Best Local Similarity 27.1%; Pred. No. 1.4e-27;  
 Matches 133; Conservative 103; Mismatches 221; Indels 33; Gaps 11;

QY 43 EPRSLDPRQVRLISELISLVKHYGLVQENNLGNIPEALAEYSLSDGLTYFFKLK- 101  
 Db 37 EPQTQQAHTSINIEFILLKLYEGLT-YYDAAGKIVPGAAETNELSDGTVYTFKLAD 95  
 QY 102 AFWSNGDPTAEDFTESWKQVATQEVSGIYAFALNP IKNRKIORGHLSIDHGVHSPNE 161  
 Db 96 AKWSDGSPVTAEDFAFSRRVEDPKTAAEYANILPIKNAEKVNKGEVPDQLGVKXAVDE 155  
 QY 162 STLWVLTESPTSHFLKLLALVPFFVHKSQ-----RTLOSKSLPIASGAFYPNIKQKW 216  
 Db 156 KTLVTLERPTFFLELLAHQTALPVSKASVEKNGADFVKPCVMVSNAGFKLTAHYPNDS 215

QY 217 IKLSKNPHVYNOSOVERKTTITHIFPDANTAAKLFNKGKLNWOGPPWGERIPQETLSNLQ 276  
 Db 216 LTVKNTINWDAAANVKLDKVIYFIPIDQQAASVRRFEAKEMD-----LAYNFSADQIESLR 270  
 QY 277 SK--CHLHSFDVAGTSMITFNINKEPLNNMKLREALASALDKEALYSTIFLGRKATADHL 334  
 Db 271 TSYGROVHVSPTLATYYAFDTROPEYNDVRRALSMVDRDFLAKEIYSGQLPSYSM 330  
 QY 335 LPTNIHSYPHQKQEMA-----ORQAYAKKLFKEALEELQITAKDLHLNLIFFPVSSAS 389  
 Db 331 VPPGIESYGDPAKADPADMSQLDREDKAIELMKEA-----GYGEGKPLNIEIRYNTNPN 385  
 QY 390 --SLLVQLIREQWKESLGFAPV-----VCKEFAFALLQADLSSGNFSLATGGWADPADMAF 444  
 Db 386 HERVATAVADWKNITGAKVSLVNLVDVSSHVAYLQ---EGGKFNVARAGWADYADAENF 442  
 QY 445 LTIFAYPSGVPPYA-INHKDFLEILQNEQDQHKRSELVSQASLYLETHIIEPIYHD 503  
 Db 443 LALSLSNTKTFNYGHFENAEFDALMKKSYBEQDPAARSKIMHEAETLLMKEQPIAPFLTQ 502  
 QY 504 AFQFAMNKKL 513  
 Db 503 ADMLVSEV 512

Search completed: July 26, 2002, 04:41:41  
 Job time: 319 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2002, 03:36:37 ; Search time 50.88 seconds

(without alignments)  
1004.708 Million cell updates/sec

Title: US-09-824-567-2

Sequence: 1 MKRISVGCITITLLSYVL.....LSNLGVSPGVDERAKEN 532

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 segs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Query	Match Length	ID	Description
1	2739	100.0	532	2 A72107	peptide ABC transp
2	2739	100.0	532	2 C86515	oligopeptide bindi
3	1364	49.8	518	2 B71545	probable oligopept
4	1345	49.1	520	2 B81699	peptide ABC transp
5	1160.5	42.4	527	2 B72107	peptide ABC transp
6	1160.5	42.4	527	2 D86515	oligopeptide bindi
7	755.5	27.6	526	2 D72107	oligopeptide bindi
8	755.5	27.6	526	2 F86515	oligopeptide bindi
9	755.5	27.6	526	2 A81564	peptide ABC transp
10	748	27.3	527	2 H81701	peptide ABC transp
11	685.5	25.0	529	2 F71547	probable oligopept
12	617.5	22.5	545	2 A38447	oligopeptide ABC t
13	536.5	19.6	550	2 B97346	oligopeptide ABC t
14	505.5	18.5	528	2 H70140	oligopeptide ABC t
15	489.5	17.9	549	2 G69618	oligopeptide ABC t
16	482.5	17.6	435	2 E86515	oligopeptide ABC t
17	482.5	17.6	435	2 C72107	oligopeptide ABC t
18	479.5	17.5	532	2 E95949	peptide ABC transp
19	478	17.5	559	2 A81719	probable oligopept
20	474.5	17.3	553	2 B98207	phorbome ABC tran
21	474.5	17.3	553	2 AG3079	hypothetical prote
22	474.5	17.3	558	2 AD1349	phorbome ABC tran
23	472	17.2	550	2 G86667	hypothetical prote
24	471.5	17.2	530	2 D70218	oligopeptide ABC t
25	467	17.1	545	2 H86667	hypothetical prote
26	446.5	16.3	553	2 AD1771	diptepide ABC tran
27	446	16.3	565	2 H97345	oligopeptide ABC t
28	438.5	16.0	544	2 D64882	periplasmic oligop
29	438.5	16.0	553	2 A11395	diptepide ABC tran

30	437.5	16.0	544	2 G90867	probable transport
31	437.5	16.0	544	2 B85751	probable transport
32	431	15.7	543	2 B82242	oligopeptide ABC t
33	428.5	15.6	529	2 B70211	oligopeptide ABC t
34	425.5	15.5	529	2 G70140	oligopeptide ABC t
35	421	15.4	549	2 A11457	ABC transporter ol
36	419	15.3	426	2 D71552	probable periplasm
37	418	15.2	538	2 AD0285	periplasmic oligop
38	417	15.2	545	2 A10285	oligopeptide-bindt
39	415	15.2	543	1 F64871	oligopeptide ABC t
40	414.5	15.1	541	2 A70141	peptide ABC transp
41	412	15.0	428	2 A81705	hypothetical prote
42	412	15.0	543	2 G90846	hypothetical prote
43	412	15.0	543	2 F85704	peptide-binding pr
44	412	15.0	549	2 C86592	periplasmic murein
45	411.5	15.0	537	2 AB0660	

## ALIGNMENTS

RESULT 1  
A72107  
peptide ABC transporter, periplasmic peptide-binding protein, probable CP0572 [import  
C:/Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C:/Date: 23-Apr-1999 #sequence-revision 23-Apr-1999 #text-change 11-May-2000  
C:/Accession: A72107, A81562  
R:/Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,  
Nature Genet. 21, 385-389, 1999  
A:/Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:/Reference number: A72000; MUID:99206606  
A:/Accession: A72107  
A:/Molecule type: DNA  
A:/Residues: 1-532 <ARN>  
A:/Cross-references: GB:AE001606; GB:AE001363; NID:94376464; PIDN:AA018348.1; PID:9437  
A:/Experimental source: strain CW029  
R:/Head, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
C:/Dodson, R.; Gwin, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salde  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:/Title: Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR35  
A:/Reference number: A81500; MUID:20150255  
A:/Accession: A81562  
A:/Molecule type: DNA  
A:/Residues: 1-532 <REA>  
A:/Cross-references: GB:AE002216; GB:AE002161; NID:97189484; PIDN:AA038391.1; PID:9716  
A:/Experimental source: strain AR39, HL cells  
C:/Genetics:  
A:/Gene: oppA\_1; CP0572  
C:/Superfamily: diptepide transport protein

Query Match 100.0%; Score 2739; DB 2; Length 532;  
Best local Similarity 100.0%; Pred. No. 1.4e-187;  
Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRISVGCITITLLSYVLQCKESSHSSRGEALNIREPPSLDPQVRLSEISL 60  
DB 1 MKRISVGCITITLLSYVLQCKESSHSSRGEALNIREPPSLDPQVRLSEISL 60  
QY 61 VKHLYGVLQENNLGNLEPALADYSLSSDGLTYFEKLSAFWSNGDPLTAEDFIESMK 120  
DB 61 VKHLYGVLQENNLGNLEPALADYSLSSDGLTYFEKLSAFWSNGDPLTAEDFIESMK 120  
QY 121 QVATQVSVITAFALNPILKNVKTQEGHSIDHFGVHSPNESTLVLTLESPSHFKLIA 180  
DB 121 QVATQVSVITAFALNPILKNVKTQEGHSIDHFGVHSPNESTLVLTLESPSHFKLIA 180  
QY 181 LPVFFVHRSQRTQSKSLPLASGAFYKNIKOKWILSNPHYVQSOVERKTTTHF 240  
DB 181 LPVFFVHRSQRTQSKSLPLASGAFYKNIKOKWILSNPHYVQSOVERKTTTHF 240  
QY 241 IPDANTAAKLINQGLNMQGPWGERIPQETLSNQSQGHSHSDVAGTSMLTNNIKFP 300  
DB 241 IPDANTAAKLINQGLNMQGPWGERIPQETLSNQSQGHSHSDVAGTSMLTNNIKFP 300

Db 241 IPDANTAKLNFQGNKLNQGPWGGERIPQETLSNLSQKGLHSPDVAGTSMLEFNINPKFP 300  
 QY 301 LNNMKLRALASALDKREALVSTIFLGRAKTADHLPTNINHSYPHOKQOEAQOAVAKKL 360  
 Db 301 LNNMKLRALASALDKREALVSTIFLGRAKTADHLPTNINHSYPHOKQOEAQOAVAKKL 360  
 QY 361 FKALEELQITAKDLEHLNLFYVSSASSLLVOLIREQKESLGFAIPVIGKEFALLQA 420  
 Db 361 FKALEELQITAKDLEHLNLFYVSSASSLLVOLIREQKESLGFAIPVIGKEFALLQA 420  
 QY 421 DLSGNFSLATGGMFADFPADPMALFIFAYPSGVPYAINNKHDFELTQNIQEQDQHR 480  
 Db 421 DLSGNFSLATGGMFADFPADPMALFIFAYPSGVPYAINNKHDFELTQNIQEQDQHR 480  
 QY 481 SELVQASLYLETFHIEPIYHDAFQFAMNKKLSNLGVSPGVDFRYAKEN 532  
 Db 481 SELVQASLYLETFHIEPIYHDAFQFAMNKKLSNLGVSPGVDFRYAKEN 532

## RESULT 2

oligopeptide binding protein [imported] - Chlamydia pneumoniae (strain J138)  
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
 C:Accession: C86515  
 R:Shirai, M.; Hiraoka, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is  
 Nucleic Acids Res. 28, 2311-2314, 2000  
 A>Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
 A:Reference number: A86491; MUID:20330349  
 A:Accession: C86515  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-532 <STO>  
 A:Cross-references: GB:BA000008; NID:98978568; PIDN:BA098405.1; GSPDB:GN00142  
 A:Experimental source: strain J138  
 C:Genetics:  
 A:Gene: oppa\_1

Query Match 100.0%; Score 2739; DB 2; Length 532;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-187;  
 Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRISVGICITLLISLVLOGCKESSHSTSGELAINIRBEPRLDPROYRLISL 60  
 Db 1 MKRISVGICITLLISLVLOGCKESSHSTSGELAINIRBEPRLDPROYRLISL 60  
 QY 61 VKHLYBGLVOENNLGNIEPALADYSLSDGLTYTFKLSAFWNGDPLTAEDFTESWK 120  
 Db 61 VKHLYBGLVOENNLGNIEPALADYSLSDGLTYTFKLSAFWNGDPLTAEDFTESWK 120  
 QY 121 QVATQVSGITAFALNPINRKTKIOEGHLSIDHFGVHSPNESTLVLTLESTSHFLKLA 180  
 Db 121 QVATQVSGITAFALNPINRKTKIOEGHLSIDHFGVHSPNESTLVLTLESTSHFLKLA 180  
 QY 181 LPVFPVHKSORTLOSLSPLASGAFYPPKNIKOKOMIKLSNPHYVNOSEVETKTYTIF 240  
 Db 181 LPVFPVHKSORTLOSLSPLASGAFYPPKNIKOKOMIKLSNPHYVNOSEVETKTYTIF 240  
 QY 241 IPDANTAKLENOGKLMWGPWGERIPQETLSNLSQKGLHSPDVAGTSMLEFNINPKFP 300  
 Db 241 IPDANTAKLENOGKLMWGPWGERIPQETLSNLSQKGLHSPDVAGTSMLEFNINPKFP 300  
 QY 301 LNNMKLRALASALDKREALVSTIFLGRAKTADHLPTNINHSYPHOKQOEAQOAVAKKL 360  
 Db 301 LNNMKLRALASALDKREALVSTIFLGRAKTADHLPTNINHSYPHOKQOEAQOAVAKKL 360  
 QY 361 FKALEELQITAKDLEHLNLFYVSSASSLLVOLIREQKESLGFAIPVIGKEFALLQA 420  
 Db 361 FKALEELQITAKDLEHLNLFYVSSASSLLVOLIREQKESLGFAIPVIGKEFALLQA 420  
 QY 421 DLSGNFSLATGGMFADFPADPMALFIFAYPSGVPYAINNKHDFELTQNIQEQDQHR 480  
 Db 421 DLSGNFSLATGGMFADFPADPMALFIFAYPSGVPYAINNKHDFELTQNIQEQDQHR 480

Db 421 DLSGNFSLATGGMFADFPADPMALFIFAYPSGVPYAINNKHDFELTQNIQEQDQHR 480  
 QY 481 SELVQASLYLETFHIEPIYHDAFQFAMNKKLSNLGVSPGVDFRYAKEN 532  
 Db 481 SELVQASLYLETFHIEPIYHDAFQFAMNKKLSNLGVSPGVDFRYAKEN 532

## RESULT 3

probable oligopeptide binding protein - Chlamydia trachomatis (serotype D, strain W  
 B71545  
 C:Species: Chlamydia trachomatis  
 C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 21-Jan-2000  
 C:Accession: B71545  
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Arevalo, L.; Mital  
 Science 282, 754-759, 1998  
 A>Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia  
 A:Reference number: A71570; MUID:99000809  
 A:Accession: B71545  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-518 <ARN>  
 A:Cross-references: GB:AE001293; GB:AE001273; NID:93328597; PIDN:AAC67790.1; PID:93;  
 A:Experimental source: serotype D, strain UW-3/Cx  
 C:Genetics:  
 A:Gene: oppa\_3  
 C:Superfamily: dipeptide transport protein

Query Match 49.8%; Score 1364; DB 2; Length 518;  
 Best Local Similarity 50.1%; Pred. No. 1.8e-89;  
 Matches 269; Conservative 97; Mismatches 147; Indels 24; Gaps 8;

QY 1 MKRISVGICITLLISLVLOGCKESSHSTSGELAINIRBEPRLDPROYRLIS 56  
 Db 1 MKRISVGICITLLISLVLOGCKESSHSTSGELAINIRBEPRLDPROYRLIS 56  
 QY 57 EISLVKHYEGLVOENNLGNIEPALADYSLSDGLTYTFKLSAFWNGDPLTAEDFT 116  
 Db 57 EISLVKHYEGLVOENNLGNIEPALADYSLSDGLTYTFKLSAFWNGDPLTAEDFT 116  
 QY 117 ESMKVQATQVSGITAFALNPINRKTKIOEGHLSIDHFGVHSPNESTLVLTLESTSHFL 176  
 Db 117 ESMKVQATQVSGITAFALNPINRKTKIOEGHLSIDHFGVHSPNESTLVLTLESTSHFL 176  
 QY 177 KILALPVFPVH-KSQRTLOSLSPLASGAFYPPKNIKOKOMIKLSNPHYVNOSEVETK 235  
 Db 177 KILALPVFPVH-KSQRTLOSLSPLASGAFYPPKNIKOKOMIKLSNPHYVNOSEVETK 235  
 QY 236 ITHIFIDANTAKLENOGKLMWGPWGERIPQETLSNLSQKGLHSPDVAGTSMLEFN 295  
 Db 236 ITHIFIDANTAKLENOGKLMWGPWGERIPQETLSNLSQKGLHSPDVAGTSMLEFN 295  
 QY 296 INKFLPNNMKLRALASALDKREALVSTIFLGRAKTADHLPTNINHSYPHOKQOEAQO 355  
 Db 296 INKFLPNNMKLRALASALDKREALVSTIFLGRAKTADHLPTNINHSYPHOKQOEAQO 355  
 QY 356 YAKLFKALEELQITAKDLEHLNLFYVSSASSLLVOLIREQKESLGFAIPVIGKEF 415  
 Db 356 YAKLFKALEELQITAKDLEHLNLFYVSSASSLLVOLIREQKESLGFAIPVIGKEF 415  
 QY 416 ALLQADLSGNFSLATGGMFADFPADPMALFIFAYPSGVPYAINNKHDFELTQNIQEQD 475  
 Db 416 ALLQADLSGNFSLATGGMFADFPADPMALFIFAYPSGVPYAINNKHDFELTQNIQEQD 475  
 QY 476 DHQKSELVQASLYLETFHIEPIYHDAFQFAMNKKLSNLGVSPGVDFRYAKEN 532  
 Db 476 DHQKSELVQASLYLETFHIEPIYHDAFQFAMNKKLSNLGVSPGVDFRYAKEN 532  
 QY 532 NPKRSALSELSEASLYIERQNVIEPLHYDHVHYHTNNKLSFVPLHSGLVDMYRAKNS 518  
 Db 532 NPKRSALSELSEASLYIERQNVIEPLHYDHVHYHTNNKLSFVPLHSGLVDMYRAKNS 518

## RESULT 4

8B1699  
 E81699  
 C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn

C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 02-Jun-2000  
 C/Accession: E81699  
 R/Read: T.D.: Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
 Nucleic Acids Res. 28, 1397-1406, 2000  
 A>Title: Genome sequences of *Chlamydia trachomatis* Mopn and *Chlamydia pneumoniae* AR39.  
 A/Reference number: A81500; MUID:20150255  
 A/Accession: E81699  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-520 <TEXT>  
 A/Cross-references: GB:AE002315; GB:AE002160; NID:q7190506; PIDN:AAF39319.1; PID:q7190517  
 A/Experimental source: strain Nigg (Mopn)  
 C/Genetics:  
 A/Gene: TC0471  
 C/Superfamily: dipeptide transport protein

```

Query Match          49.1%; Score 1345; DB 2; Length 520;
Best Local Similarity 43.4%; Pred. No. 46-88;
Matches 266; Conservative 103; Mismatches 145; Indels 24; Gaps 9;

OY 1 MRKISVGIITLLSLVVLQCKESS-----HSTRSGELAINRDEPSIDPQVRL 55
DB 1 MRKVSYGIC--LLIMATATIGCKSSSKNSNNSSNO-SYVSXKDDPRTDPREVRL 57
OY 56 SEISLVKHYEGLVOENNLNLEPDAEDYSLSDGLTYTFKLSAFWSNGDPLTAEDF 115
DB 58 SDINLHLHYELVQETP-SGEVFPALAESEKTKTYFHLKALMNGDLIAHNF 116
OY 116 IESMKOVATQEVSGIYAFALNPKIKNRKIQEGHLSIDHFGVHSNRETLVLTSEPTSHF 175
DB 117 VRSMNDVQGNRIASISYFAFLPT-----DLSKD-CGFPAQDNHTLVINLHPTPIHF 166
OY 176 LKILALPVPFVYH-KSQPTLQSKSLPIASGAFYFNKIKOKWIKLSKNPHYVQSOVERX 234
DB 167 LKLLLPVFPYVEHQIRNEAKALPISTGAFCLKEKKDRKMLKLENPITYKKEVALIQ 226
OY 235 TITIFHPDANNAKLENOGKLMQCPWPVEGRIPOETLSNLSQKGLHSPDVAGTSWLT 294
DB 227 ELIHIVIPDQATLAFKQKGLMQLPQGHSHIQEALATANKRTPQSPDISGTSWLT 286
OY 295 NINKEPFLNMKLEALASALDKELAVSTFTLGRAKTADHLLPNIHSYEPHOKOEAQRQ 354
DB 287 NTSKVPFSHPKLRQALSLVNLKALASPTFV---KPAKHLPLHLHYEQPTTYQOBAV 343
OY 355 AYAALKEALEELQITAKDLLEHLNLPVSSASSLLVQLEIOWKESLGFAPIVGKE 414
DB 344 ILAKTLQEAULTDLMNTIDLEKCPILFSATSSVNSQMAQMDQRRLTIGITPICK 403
OY 415 FALLOADLSSGNFSLATGFWFADFPADPMAFLITAFYSSGVPFALNHKDFLELONIQE 474
DB 404 YALLQNDLTNTTFKSIINGWPADESDPLAFLVSFS-SKGIKPYALODPLDLILSTITE 462
OY 475 ODHOKRSLEYQASLYLETFHHIEPIYHDAFOFAMNKKLSNLGVSPTGVFRRAKEN 532
DB 463 KDPKRKISLISEASTYLEKQNTIEPLHYHDAVHYAANNKLSFVALHSGLVMDRIAKNS 520

```

RESULT 5  
 B72107

peptide ABC transporter, periplasmic peptide-binding protein, probable CP0571 [imported]  
 C/Species: *Chlamydia pneumoniae*, *Chlamydia pneumoniae*  
 C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
 C/Accession: B72107; C81564  
 R/Katman, S.; Mitchell, W.; Marathe, R.; Tamm, J.; Olinger, L.; Grimwood, J.;  
 Nature Genet. 21, 385-389, 1999  
 A>Title: Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*.  
 A/Reference number: AY2000; MUID:9920606  
 A/Accession: B72107  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-527 <ARN>

A/Cross-references: GB:AE001606; GB:AE001363; NID:q4376464; PIDN:AD18349.1; PID:q437  
 A/Experimental source: strain CW1029  
 R/Read: T.D.: Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
 Nucleic Acids Res. 28, 1397-1406, 2000  
 A>Title: Genome sequences of *Chlamydia trachomatis* Mopn and *Chlamydia pneumoniae* AR39.  
 A/Reference number: A81500; MUID:20150255  
 A/Accession: C81564  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-527 <REA>  
 A/Cross-references: GB:AE002215; GB:AE002161; NID:q7189472; PIDN:AAF38390.1; PID:q718  
 A/Experimental source: strain AR39, HL cells  
 C/Genetics:  
 A/Gene: opa\_2; CP0571  
 C/Superfamily: dipeptide transport protein

```

Query Match          42.4%; Score 1160.5; DB 2; Length 527;
Best Local Similarity 43.2%; Pred. No. 66-75;
Matches 225; Conservative 111; Mismatches 176; Indels 9; Gaps 4;

OY 15 SLVVLQCKESSHSTRSGELAINRDEPSIDPQVRLSLISYKHYEGVQENL 74
DB 11 TLMLITSGC---SPQSSKGIQVYMAKMPRLSDPKTRLIAQYTLRHYEGVLSHQ 67
OY 75 SGNEPLADYSLSDGLTYTFKLSAFWSNGDPLTAEDFLESWKOVATQEVGIYAF 134
DB 68 NGELKIPALASYSIISDEGTREYFKIKNLWNSGDLPLAQDFVSSKKEIKEDASVLYLA 127
OY 135 LNPKNYKIQEGHLSIDHFGVHSNRETLVLTSEPTSHFLLALPVPFVHKSQRT 194
DB 128 FLPIKARALPDTQSPENLVGVALDKRHLEIQLEPCAHFLPFLTPIFPVHETLRNY 187
OY 195 QS--KSLPIASGAFYFNKIKOKWIKLSKNPHYVQSOVERXITITIFHPDANNAK 252
DB 188 STSPFEMPITCGAFRPVSLKGLALHLEKMPATHNSRKILKTIIVDSIVANALILRK 247
OY 253 QGKLMQCPWPVEGRIPOETLSNLSQKGLHSPDVAGTSWLFNINKEPFLNMKLEALAS 312
DB 248 HKKLDQGPWPVEGRIPOETLSNLSQKGLHSPDVAGTSWLFNINKEPFLNMKLEALAS 307
OY 313 ALDKELAVSTFTLGRAKTADHLLPNIH--SYEPHOKOEAQYAKLKEALELQIT 370
DB 308 AIDKMLTVVYQGLAETDHLHPLRYETPYERRKQV--ERILEAQOQFEELADELOM 365
OY 371 TANDLEHLNLPVSSASSLLVQLEIOWKESLGFAPIVGKEFALLODSSGNFSLA 430
DB 366 TREDEKEKTLIFSTFSFGYRICQLRQMKVYLFTPIYVQGFETIQNLEGNYSLT 425
OY 431 TCGWFADFPADPMAFLITAFYSSGVPFALNHKDFLELONIQEBOBQKRSLEYQASLY 490
DB 426 VQWMTAAFTDPSYMTKIPANFGISPHQDSHFQTLIKITQKHKLNLQILIALDY 485
OY 491 LETFHIEPIYHDAFOFAMNKKLSNLGVSPTGVFRRAKE 531
DB 486 LEHCHTLEPLCHPMLRALKNKIKNPLNLFVKTSPFRIEK 526

```

RESULT 6  
 D86515

oligopeptide binding protein [imported] - *Chlamydia pneumoniae* (strain J138)  
 C/Species: *Chlamydia pneumoniae*, *Chlamydia pneumoniae*  
 C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
 C/Accession: D86515  
 R/Shital, M.; Hirkawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;  
 Nucleic Acids Res. 28, 2311-2314, 2000  
 A>Title: Comparison of whole genome sequences of *Chlamydia pneumoniae* J138.  
 A/Reference number: A86491; MUID:20330349  
 A/Accession: D86515  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-527 <STO>

1. The first step in the process of identifying a problem is to recognize that a problem exists. This involves gathering information about the situation and identifying the specific issue that needs to be addressed.

```

Db 192 ESKKQTPSTYISNGPVLKKHSHQNYLLEKNPHYHDSVSKLDRTLTILIPDSTAT 251
QY 249 KLFNOGKLMWGPGERIPQETLSNLSQSGHSHSDVAGSWLTPEINKEPLNNKLE 308
Db 252 KLFKSKSIDWISPMASPINED-QAVLSQEKILTYSSVSTLLIYNLQKPLIONKALK 310
QY 309 ALASALDKKALVSTIFLGAKTADHLPTNHSYPEHQKQMAORQAVAKILKEALEL 368
Db 311 ALHAIDRKSTILRLVPSG--QEAVLTPPNLSQNLQKEISTEERQTKAKATFOEAKETL 368
QY 369 QITAKDLEHLNLFPPVSSASSLLVOLIREQKESLGAIPYVGEKFPALLQADLSSGNFS 428
Db 369 --SEKLAELSLIPLIDSSNSIIAQEIQRQLKDTLGKIKIQGMEYHCKLRKRGDDF 426
QY 429 LATGWFADPADMAFLITFAVPSGVPYAIINHKDFLELIONT---EEDODHOKSELS 485
Db 427 IATGGMIAEVSVPALSTILGNPRDITQW--RNSDYETKLEKTYLPHAYKENLKRAEMI- 483
QY 486 QASLYLEFPHLEPIYHDAFOFAMNKKLSNLGVSPTGVVDFR 527
Db 484 -----IEEPTPIPLHGYIYAIHPKIONTFGSLGHTDLK 520

RESULT 9
A:1564
peptide ABC transporter, periplasmic peptide-binding protein CP0569 [imported] - Chlamy
C:Species: Chlamydomonas pneumoniae
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Jun-2000
C:Accession: A81564
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MOPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: A81564
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-528 <RFA>
A:Cross-references: GB:AE002215; GB:AE002161; NID:g7189472; PIDN:AAF38388.1; PID:g718948
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0569
C:Superfamily: dipeptide transport protein

```

```

Query Match 27.6%; Score 755.5; DB 2; Length 528;
Best Local Similarity 33.9%; Pred. No. 4,8e-46;
Matches 177; Conservative 106; Mismatches 216; Indels 23; Gaps 9;

```

```

QY 13 LLSLVYLOGCKESSHSTSGELAINIRDEPSLDPRQVRLSLSLVKHIYEGLYQEN 72
Db 17 LFLPLTLSSCKQOEPLGK-HLIVIMSHDLADLPPRAVYLSRNASLAKALVEGLRREI 75
QY 73 NLSGTEPALADYSLSSDGLTYTFKLKSAFNSGDPPLTAEDFIESKQAVATQVSGIYA 132
Db 76 D--QGIALALASLYLSKDHVYTFKLKPSVWSGDPPLTAEDFIESKQAVATQVSGIYA 132
QY 133 PALNPKNKRIQEGHSLIDHFGVHSPNDSLVTLESPTSHPLKILALVPEPVKSR 192
Db 134 TLLGVINKSSAIIHNKQSLFTLIGIAKDDLTLYLEDPPEYFTLLTARVESPVHHLR 193
QY 193 TLQSKSLP---IASGAFYRKNIKQKOWIKLSKNPHYVNSOVBETKTLTHFPDANTAA 248
Db 194 ESKKQTPSTYISNGPVLKKHSHQNYLLEKNPHYHDSVSKLDRTLTILIPDSTAT 253
QY 249 KLFNOGKLMWGPGERIPQETLSNLSQSGHSHSDVAGSWLTPEINKEPLNNKLE 308
Db 254 KLFKSKSIDWISPMASPINED-QAVLSQEKILTYSSVSTLLIYNLQKPLIONKALK 312
QY 309 ALASALDKKALVSTIFLGAKTADHLPTNHSYPEHQKQMAORQAVAKILKEALEL 368
Db 311 ALHAIDRKSTILRLVPSG--QEAVLTPPNLSQNLQKEISTEERQTKAKATFOEAKETL 370

```

```

QY 369 QITAKDLEHLNLFPPVSSASSLLVOLIREQKESLGAIPYVGEKFPALLQADLSSGNFS 428
Db 371 --SEKLAELSLIPLIDSSNSIIAQEIQRQLKDTLGKIKIQGMEYHCKLRKRGDDF 428
QY 429 LATGWFADPADMAFLITFAVPSGVPYAIINHKDFLELIONT---EEDODHOKSELS 485
Db 429 IATGGMIAEVSVPALSTILGNPRDITQW--RNSDYETKLEKTYLPHAYKENLKRAEMI- 485
QY 486 QASLYLEFPHLEPIYHDAFOFAMNKKLSNLGVSPTGVVDFR 527
Db 486 -----IEEPTPIPLHGYIYAIHPKIONTFGSLGHTDLK 522

```

```

RESULT 10
A:1701
peptide ABC transporter, periplasmic peptide-binding protein TC0446 [imported] - Chla
C:Species: Chlamydia muridarum, Chlamydia trachomatis MOPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Jun-2000
C:Accession: H81701
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MOPn and Chlamydia pneumoniae AR39
A:Reference number: A81500; MUID:20150255
A:Accession: H81701
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-527 <RFA>
A:Cross-references: GB:AE002313; GB:AE002160; NID:g7190484; PIDN:AAF39300.1; PID:g719
A:Experimental source: strain N199 (MOPn)
C:Genetics:
A:Gene: TC0446
C:Superfamily: dipeptide transport protein

```

```

Query Match 27.3%; Score 748; DB 2; Length 527;
Best Local Similarity 35.3%; Pred. No. 1.6e-45;
Matches 185; Conservative 96; Mismatches 217; Indels 26; Gaps 12;

```

```

QY 14 LLSLVYLOGCKESSHSTSGELAINIRDEPSLDPRQVRLSLSLVKHIYEGLYQEN 73
Db 14 LSLFTGLSLCHQKEEN--LRNITRAICHDPWSDPRQVRLSLSLVKHIYEGLYQEN 71
QY 74 LSGTEPALADYSLSSDGLTYTFKLKSAFNSGDPPLTAEDFIESKQAVATQVSGIYA 133
Db 72 --GSLHLAELRYSLSDQRCVYTFKLKTFPHNGDLYTALDEESTKQTYLAEVDNALR 129
QY 134 ALNPKNKRIQEGHSLIDHFGVHSPNDSLVTLESPTSHPLKILALVPEPVKSR 190
Db 130 LLAIKNSHAYLKGDIPEVNLGVRALIDEHTELTLEHPSHPELTTHPVFVNASLRE 189
QY 191 -ORTLQSKSLPIAS-GAFYRKNIKQKOWIKLSKNPHYVNSOVBETKTLTHFPDANTAA 248
Db 190 YRRNRKRSPLIISNGPFIIRCYEPQNTLLDKNPFYHDDKNVSLDAVLAQIVPDIHTAV 249
QY 249 KLFNOGKLMWGPGERIPQETLSNLSQSGHSHSDVAGSWLTPEINKEPLNNKLE 308
Db 250 QLFQKKYVDVLDVGMSSSPLEBQKNL-SQDFLYDLYVNLGTYLCVNNHPLNDPSLRA 308
QY 309 ALASALDKKALVSTIFLGAKTADHLPTNHSYPEHQKQMAORQAVAKILKEALEL 368
Db 309 ALSLAIDRETLK--LAGGSLATSFVAPSLSKPLDVLISQ--KERISLAKNYLAELAK-- 363
QY 369 QITAKDLEHLNLFPPVSSASSLLVOLIREQKESLGAIPYVGEKFPALLQADLSSGNFS 428
Db 364 TVROBELKKTILITPLSTIVLRAVVOELRQDLDPVAGFISYTLGLEYHFDKRSKGEFS 423
QY 429 LATGWFADPADMAFLITFAVPSGVPYAI--NKKDFLELION--EEDODHOKSELS 483
Db 424 LSTGNVADYQOAKAFSLIDNGTKYKDYIDIMQKQETDLYVSLVLESDTDLIQILAE- 482
QY 484 VSOASLYLEFPHLEPIYHDAFOFAMNKKLSNLGVSPTGVVDFR 527

```



Db 483 ----QLLKESPLI-PIVHLDYAIAPKPKVSNLQISSIGELIDK 521

## RESULT 11

F71547

Probable oligopeptide binding protein - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
C:Species: Chlamydia trachomatis  
C>Date: 13-Sep-1998 #sequence-revision 13-Sep-1998 #text-change 21-Jan-2000  
C:Accession: F71547

R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marthe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trac

A:Reference number: A71570; MUID:99000809

A:Accession: F71547

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-529 <ARN>

A:Cross-references: GB:AE001291; GB:AE001273; NID:93328573; PIDN:AC67766.1; PID:9332857

A:Experimental source: serotype D, strain UW-3/Cx

C:Genetics:

A:Gene: oppA\_2

C:Superfamily: dipeptide transport protein

## Query Match

Best Local Similarity 25.0%; Score 685.5; DB 2; Length 529;  
Matches 181; Conservative 89; Mismatches 227; Indels 29; Gaps 10;

Db 14 LSLAVYLQCKESSHSSTRGELAINIDEPSPDPQVYLSEIYLHYIEGJQENN 73

Db 14 VSLAFGLTSCYHKKE--PKDVLRIALICHDMPSLDPROVFLSDYKALIEGLVMEKE 71

Db 74 LSGNIPALAEVSYSSDGLTYTFKLSAFNSNDPLIADPTESMKQVAFQVGIYAF 133

Db 72 AA--FQIALAEFRHSDGCVTFPLKTFMSNGDVYTADEESIKOYFREINDPSLR 129

Db 134 ALNPKYKWKIOEGHLSIDHEGVSNSSTLVTLSPSHFLKIALPVPYHKSQRT 193

Db 130 SLAIKNSHNVLTGALPVEDIGVRLNAKLEIYLNPPYLETLLAHVPYVHTSLRE 189

Db 194 L-----QSKSLPIAS-GAFYKRNKQKQIKLSKNHYVNOGVEFTKITITFPDANPA 247

Db 190 YKDKRKRPFLISNGFPALQCEPQRYLLINKNPLYHAKHVDLNSVCLQIYDPLDHA 249

Db 248 AKLFNGCKLNMGGPPWGERIPQETLSLQSKGHLSEFDAGISLNTENKFPILNNKLR 307

Db 250 MQLFQKNHIDVLGIPWSSSFLSEQRNL-PREKLDYVPLSCSVLFCNIHQTFLLNPSLR 308

Db 308 EALASLQDEALVSTIFLGRAKTADELPTNHSYPERHOKQEMARQAVAKLFEKALRE 367

Db 309 TALSLINTEFLK--LAGKCSATSFVHPQLSQIPATTLGQ-DEKIALAKGILTALKT 365

Db 368 LQTTAKDELHLNLFYVSSASSLLVOLLREQMKESLFAPIYKGFALLQADSSGNF 427

Db 366 L--SQEDLEKITILYPIESVCLAVVQLRQGLFVLTGLISTGLYHGFCDKRSRGEF 423

Db 428 STATGWAADPAWAFVLTFAVPSGVPIYANKEDELEI-----LQNEDEQDHQKRS 481

Db 424 STATGWAADPAWAFVLTFAVPSGVPIYANKEDELEI-----LQNEDEQDHQKRS 481

Db 482 ELVSOASILETFPIIPIPIYHAFQPMKKLSNIGSPGVVDFR 527

Db 477 DLQMAEQDLLKESPLIPIVHLDYAIAPKPKVSNLQISSIGELIDK 522

## RESULT 12

A38447

oligopeptide ABC transport system substrate-binding protein oppA precursor - Bacillus su  
N:Alternate names: oppA homolog; sporulation initiation protein spo0A  
C:Species: Bacillus subtilis  
C>Date: 12-Jul-1991 #sequence-revision 12-Jul-1991 #text-change 20-Jun-2000  
C:Accession: A38447; S15230; G69668

R:Rudner, D.Z.; LeDeaux, J.R.; Ireton, K.; Grossman, A.D.  
J. Bacteriol. 173, 1388-1398, 1991

A:Title: The spo0K locus of Bacillus subtilis is homologous to the oligopeptide perm

A:Reference number: A38447; MUID:91139580

A:Accession: A38447

A:Molecule type: DNA

A:Residues: 1-545 <RND>

A:Cross-references: GB:M57689; NID:q143602; PIDN:AAA62687.1; PID:q143603

R:Perego, M.; Higgins, C.F.; Pearce, S.R.; Gallagher, M.P.; Hoch, J.A.  
Mol. Microbiol. 5, 173-185, 1991

A:Title: The oligopeptide transport system of Bacillus subtilis plays a role in the i

A:Reference number: S15230; MUID:91194547

A:Accession: S15230

A:Molecule type: DNA

A:Residues: 1-25, 'r', 27-194, 'E', 196-339, 'M', 341-545 <PER>

A:Cross-references: EMBL:X56347; NID:940004; PIDN:CA93787.1; PID:940005

R:Kunstl, F.; Ogasawara, N.; Moszer, I.; Albertin, A.M.; Alloni, G.; Azevedo, V.; Ber  
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehlich, S.D.; Emmer, P.T.; Ertlan, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M  
A.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Landino  
Y., M.; Ogawa, K.; Ogata, A.; Odega, B.; Park, S.H.; Parto, V.; Pohl, T.M.; Portele  
A.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scari  
A.; Authors: Scheich, S.; Schreier, R.; Scoffone, F.; Seliguchi, J.; Sekowska, A.; Se  
t.; Winters, P.; Wipat, A.; Yamamoto, T.; Yarnate, P.; Yarnate, P.; Yarnate, V.; Uchaya  
A.; Authors: Yoshikawa, H.F.; Zumsel, E.; Yoshikawa, H.; Dancho, K.; Yata, K.; Yoshida  
A.; Title: The complete genome sequence of the gram-positive bacterium Bacillus subtili

A:Reference number: A69580; MUID:98044033

A:Accession: 669668

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-545 <RND>

A:Cross-references: GB:299110; GB:AL009126; NID:92633472; PIDN:CA813000.1; PID:926334

A:Experimental source: strain 168

A:Note: (Initiation of sporulation, competence development)

C:Comment: This operon of five genes is homologous to the Oligopeptide permease opero

C:Genetics:

A:Gene: oppA; spo0A

C:Superfamily: dipeptide transport protein

C:Keywords: blocked amino end; lipoprotein; membrane protein; oligopeptide transport;

F.1-20/Domain: signal sequence #status predicted <SIC>

F.21/Binding site: sn-2,3-diacylglycerol (Cys) (covalent) #status predicted

F.21/Modified site: fatty acylated amino end (Cys) (in mature form) #status predicted

C:Genetics:

A:Gene: oppA; spo0A

C:Superfamily: dipeptide transport protein

C:Keywords: blocked amino end; lipoprotein; membrane protein; oligopeptide transport;

F.1-20/Domain: signal sequence #status predicted <SIC>



...XESTVAVYANNENIA-HAEVSPGFIQPPDGKDFREAGG 367

```

QY 352 - -OROVAAKKELELELOITAKDEHNIIPVSSASS-----LLVOLREPKKE 40%
Db 368 DLIKNESKAKQJLEKGMKE-----ENMKPLITLTITSYREPKKIAELIDQKLN 41%
QY 403 SIGFALPIVGEFRLLOALSSGNFLATGCMFADPDMALFTFAVPSGVPAVNIHK 46%
Db 420 SIGVYVKIANKEMWLELDQKLFKQFSOSSFLPDYADPISFLAEFOTGSMNIRGMANK 47%
QY 463 DELEILLONIEDDHOKRSELVSQASLYLTFHILEPIYHAGFOFAMNKILNLSVSPG 52%
Db 480 EYDOLIKANKNDEKTRFSLMHQAEELLINAPRIIPVFNQVHILANDQVAGIYRNVG 53%
QY 523 VYDERYAKEN 53%
Db 540 YIDLRKADKN 54%

```

Search completed: July 26, 2002, 04:39:59  
Job time: 3802 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2002, 03:01:42 ; Search time 62.89 Seconds  
(Without alignments) 939,597 Million cell updates/sec

Title: US-09-824-567-2

Perfect score: 2739  
Sequence: 1 MKRISVSGICITLLLSLVVL.....LSNIGVPTGVDFRYAKEN 532

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

A\_Geneseq\_032802.\*  
1: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS5/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	2739	100.0	532	22	AAE12212
2	2645	96.6	514	20	AAV34796
3	1126	41.1	494	20	AAV34797
4	864	31.5	339	20	AAV37005
5	753.5	27.5	534	20	AAV34799
6	680.5	24.8	529	20	AAV37027
7	617.5	22.5	544	21	AAV97040
8	521.5	19.0	214	20	AAV37004
9	505.5	18.5	512	20	AAV20067
10	505.5	18.5	529	20	AAV20066
11	482.5	17.6	448	20	AAV34798

12	455	16.6	547	20	AAV00030
13	452.5	16.5	528	20	AAV00031
14	425.5	15.5	506	20	AAV19870
15	421	15.4	506	20	AAV19871
16	417.5	15.2	553	20	AAV00052
17	414.5	15.1	541	20	AAV19988
18	403.5	14.7	542	20	AAV00040
19	394	14.4	550	22	AAU34906
20	393.5	14.4	521	20	AAV00041
21	390.5	14.3	514	20	AAV19989
22	390	14.2	532	20	AAV00053
23	385	14.1	522	15	AAV48035
24	371.5	13.6	310	20	AAV37060
25	356.5	13.0	546	20	AAV00050
26	351	12.8	406	20	AAV00066
27	349	12.7	526	20	AAV00067
28	329.5	12.0	373	20	AAV00051
29	321.5	11.7	372	20	AAV00060
30	308.5	11.3	381	22	AAU03659
31	303.5	11.1	342	20	AAV00061
32	301	11.0	524	22	AAU34786
33	296.5	10.8	321	20	AAV00089
34	296.5	10.8	343	20	AAV00088
35	285	10.4	537	22	AAU36431
36	273.5	10.0	535	22	AAU38208
37	270.5	9.9	643	16	AAV70152
38	268.5	9.8	549	19	AAV98830
39	268.5	9.8	549	22	AAU35720
40	264	9.6	278	22	AAV35720
41	263.5	9.6	314	18	AAV89746
42	235	8.6	556	22	AAV96074
43	214.5	7.8	503	22	AAV96074
44	210.5	7.7	705	22	AAV96074
45	209	7.6	615	22	ABG15826

#### ALIGNMENTS

RESULT	1	
AAE12212	AAE12212 standard; Protein: 532 AA.	
AC	AAE12212:	
XX	15-JAN-2002 (first entry)	
DE	Chlamydia pneumoniae ATP-binding cassette protein.	
XX	ATP-binding cassette; antibiotic; vaccine; infection; therapy; poxvirus.	
KW	Chlamydia pneumoniae.	
OS	Chlamydia pneumoniae.	
XX		
PH	Key	Location/Qualifiers
FT	Region	53..61
FT	Region	/label= T-cell epitope
FT	Region	338..359
FT	Region	/label= B-cell epitope
FT	Region	469..482
FT	Region	/label= B-cell epitope
XX		
PN	WO200174863-A2.	
XX		
PD	11-OCT-2001.	
XX		
PR	04-APR-2001; 2001WO-CA00455.	
XX		
PR	04-APR-2000; 2000US-194464P.	
XX		
PA	(AVER ) AVENTIS PASTEUR LTD.	
XX		
PI	Mordin AD, Oomen RP, Wang J, Dunn P;	
XX		

Enterococcus faeca  
Enterococcus faeca  
B. burgdorferi ant  
B. burgdorferi ant  
Enterococcus faeca  
B. burgdorferi ant  
Enterococcus faeca  
Enterococcus faeca  
Enterococcus faeca  
B. burgdorferi ant  
Enterococcus faeca  
Hyaluronic acid sy  
Chlamydia trachoma  
Enterococcus faeca  
Enterococcus faeca  
Enterococcus faeca  
Enterococcus faeca  
Enterococcus faeca  
Group B Streptococ  
Enterococcus faeca  
E. coli cellular p  
Enterococcus faeca  
Pseudomonas aerugi  
Salmonella typhi c  
Streptococcus pneu  
H. pylori GPO 149  
Helicobacter pylori  
S. epidermidis ope  
Staphylococcus aur  
putative P. abyssal  
C glutamicum prote  
putative P. abyssal  
Novel human diagno

DR WPI: 2001-648549/74.  
 DR N-PSDB: AAD20238.  
 PT Novel Chlamydia ATP-binding cassette and corresponding DNA molecule for  
 PT preventing, diagnosing and treating Chlamydia infections in mammals, in  
 PT particular humans -  
 PS  
 PS Claim 21; Fig 1; 88pp; English.  
 CC The present invention relates to novel Chlamydia pneumoniae ATP-binding  
 CC cassette protein and its corresponding gene. Sequences of the invention  
 CC are useful for detecting Chlamydia infection by assaying a body fluid  
 CC of a mammal with the components. They are also used as vaccines. ATP  
 CC binding cassette antibodies and vaccines of the invention are useful  
 CC for preventing or treating Chlamydia infection e.g. infection caused  
 CC by *C. trachomatis*, *C. psittaci*, *C. pneumoniae* or *C. pecorum* in mammals,  
 CC such as humans. The nucleic acid molecules are useful for producing  
 CC ATP-binding cassettes, in the construction of vaccine vectors such  
 CC as poxviruses, which are further useful for preventing and/or treating  
 CC Chlamydia infection and in the construction of attenuated Chlamydia  
 CC strains that can over-express the nucleic acid molecules or express  
 CC it in a non-toxic, mutated form. The present sequence is ATP-binding  
 CC cassette from Chlamydia pneumoniae.  
 XX  
 XX Sequence 532 AA:

Query Match 100.0%; Score 2739; DB 22; Length 532;  
 Best Local Similarity 100.0%; Pred. No. 2,7e-240;  
 Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MRKISVGCITITLLISVYVLOCKKSSHSSTSRGELAINIRDEPSLDPROYLLSISL 60
DB 1 mrksvgtictllslsvyvgcksshsstsrgealnrdeprslprgyvllseisl 60
QY 61 VKHIYEGVQENNTSGNIEPALAEDYSLSGGLTYFFKLSAFWNGSPDLAEDFISWK 120
DB 61 vkhiyeglvqenmlsgniepalaedyssdgltlyffklsafwngspdlaedfieswk 120
QY 121 QVATQEVSGIYAFALNPTRKTRKIOEGHLSIDHFGVHSPNESTLVLTLESPTSHFLKLA 180
DB 121 qvatqevsgiyafalnptrktrkioeghlsidhfgvhspsnestlvltlespshflklla 180
QY 181 LPVFFPVHKSORTLOSKSLPIASGAFYPRKIKOKIKIRSKNPHRYNOSOVERKTTTTF 240
DB 181 lpfvfpvhsqrltqskslpiasgafyprkikqkwiksknphryngsveckttltinf 240
QY 241 IPDANTAAKLFNQGKLMQGPWGERIPQETLSNLSQKHLSFSPVAGTSLTFNNKRP 300
DB 241 ipdantaaklfngklnwgppwgeripqetlsnlsqkhlsfsvagtsltnlnktr 300
QY 301 LNNMKIREALASALDKKALVSTIFLGRATADHLLPTNHSYPRHOKQENAGQAVAKKL 360
DB 301 lnnmkirealasaalddkalvstiflgraktadhllptnhsyprhokqenagqavakkl 360
QY 361 FKRALEELQITAKDLEHNLIFPVSSASSLLVOLLIRQWKESEGFIPIVGEFALLQA 420
DB 361 fkealeeqitakdehnlifpvssassllvqllirqwkeseigfiipivgefella 420
QY 421 DLSNGFSLATGWFADPMAFLTFAPYPSGPPYALINHKDPLETIOIEEDOHKR 480
DB 421 dlsngfslatgwfadpmaflltfaypsgppyalinhkdpletioieedohkr 480
QY 481 SELVSOASILETHIIEPIYHDAFOPANKKLSLGVSPGVVDERYAKEN 532
DB 481 selvsoasilethiiepiyndaqfamnklslgvspgvvdfiyeken 532

```

RESULT 2  
 AAY34796  
 ID AAY34796 standard; Protein: 514 AA.  
 AC AAY34796;

XX 13-SEP-1999 (first entry)  
 XX Chlamydia pneumoniae lipoprotein sequence.  
 DE  
 DE Chlamydia pneumoniae; bronchitis; heart disease; sarcoidosis;  
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
 KW vaccine; neutralising epitope.  
 XX  
 OS Chlamydia pneumoniae.  
 XX  
 PN MO9927105-A2.  
 XX  
 XX 03-JUN-1999.  
 XX  
 XX 20-NOV-1998; 98WO-IB01890.  
 XX  
 PR 04-NOV-1998; 98US-0107078.  
 PR 21-NOV-1997; 97FR-0014673.  
 XX  
 PA (BEST ) GENSET.  
 XX  
 PT Griffiths R;  
 XX  
 DR WPI: 1999-357842/30.  
 XX  
 PS Genome sequence of Chlamydia pneumoniae  
 PS Page 769-770; Disclosure; 1912pp; English.

AAY34584-735879 represent the proteins encoded by all the open reading  
 CC frames in the complete genome (see AAY31990) of Chlamydia pneumoniae.  
 CC C. pneumoniae causes respiratory disease such as pneumonia and  
 CC bronchitis and is thought to be a contributing factor in heart  
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
 CC frames of the C. pneumoniae genome (see AAY34584-735879) can be used in  
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
 CC nucleotide sequences can also be used as immunogenic compositions,  
 CC especially where the vector directs the expression of a neutralising  
 CC epitope of C. pneumoniae.  
 XX

SQ Sequence 514 AA:

Query Match 96.6%; Score 2645; DB 20; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 9e-232;  
 Matches 514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MRKISVGCITITLLISVYVLOCKKSSHSSTSRGELAINIRDEPSLDPROYLLSISL 60
DB 1 mrksvgtictllslsvyvgcksshsstsrgealnrdeprslprgyvllseisl 60
QY 61 VKHIYEGVQENNTSGNIEPALAEDYSLSGGLTYFFKLSAFWNGSPDLAEDFISWK 120
DB 61 vkhiyeglvqenmlsgniepalaedyssdgltlyffklsafwngspdlaedfieswk 120
QY 121 QVATQEVSGIYAFALNPTRKTRKIOEGHLSIDHFGVHSPNESTLVLTLESPTSHFLKLA 180
DB 121 qvatqevsgiyafalnptrktrkioeghlsidhfgvhspsnestlvltlespshflklla 180
QY 181 LPVFFPVHKSORTLOSKSLPIASGAFYPRKIKOKIKIRSKNPHRYNOSOVERKTTTTF 240
DB 181 lpfvfpvhsqrltqskslpiasgafyprkikqkwiksknphryngsveckttltinf 240
QY 241 IPDANTAAKLFNQGKLMQGPWGERIPQETLSNLSQKHLSFSPVAGTSLTFNNKRP 300
DB 241 ipdantaaklfngklnwgppwgeripqetlsnlsqkhlsfsvagtsltnlnktr 300
QY 301 LNNMKIREALASALDKKALVSTIFLGRATADHLLPTNHSYPRHOKQENAGQAVAKKL 360
DB 301 lnnmkirealasaalddkalvstiflgraktadhllptnhsyprhokqenagqavakkl 360

```

```
Query Match      41.18; Score 1126; DB 20; Length 494;
Best Local Similarity 43.8%; Pred. No. 1.5e-33;
Matches 217; Conservative 104; Mismatches 168; InDels 6; Gaps 3;

QY   41 RDPPSLDPROYRLSTLSLVKHIYEGVLQSNNTSGNIEPALADYSTDSGDTYTRIK 100
      ::|||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db    1 kempisrlpdygrrlradqimrlnlyegylveehsqngsklpalesytlseagtyllkik 60
      : |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY   101 SAFWMSNPDLAEPIFSMKVQAIVESGIAPALINDKNVRKILOGHSLIDHGVSFNP 160
       : |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

D	b6	nlhwngopbltaqdfvsskkelkedasavylafipiknaraiiddtespenlgyrald	120
Oy	161	ESTLWVLESTPHEFLKLALPEFPVPKHSORTLOS--KSPLIASGAFYPRNIKIKOMWK	218
D	121	khlleiqfepcaphihltplffpphehrlrygstseempicgdtfrpslekglilh	180
Oy	219	LKKNHYYNCSOYEETITRIHTEPDANTAAKLFNQCKLMWGCPWGERIPQETLSLNQSK	278
D	181	Iekpmymnksrvklhkllvgflshantaalifkhkkldwgpdpwgepdlppeisaalhqd	240
Oy	279	GHLISFVAGTASWLTPNIKEPFLNNMKTREALASALDKREALVSTIILGRAKTKADHLFTN	338
D	241	dqlfsldgsatcttlnfnickprwmaklrassaldxamltkvvyygaaleptdhilnp	300
Oy	333	TH--SYEPHKOKMAPOAYAVAKLFKEALELOITAKDLHLNLFFVSSSASSLLVOLi	396
D	301	lytpfyetkrqn--erlleaqgfleaaldqmredtelktfstsfsgylicml	358
Oy	397	REOKESLGAPIDPYGKEPFALLQDLSGGNFSLATCGMFADFDPMAFLTIRAYSGVP	456
D	359	regharkykltktpivggelftgknfglegysltvnqtaalfdmasyimlfampgsisp	418
Oy	457	VAINHKOFEILLONIEODQHQRSELTVSQSASYLETFRHITEPIYHADAFOPANKXLSNL	516
D	419	yhlqdsfhfgyllikkgtgkhhlnqlileadyehchlepichpnrliraixnkxf	478
Oy	517	GVSPGTGVDERAKE	531
D	479	nlfvrtsdrfkke	493
RESULT	4		
AAB37005	ID	AAB37005 standard; Protein: 339 AA.	
AC	AAB37005:		
DT	07-OCT-1999	(first entry)	
DE	Chlamydia trachomatis lipoprotein sequence.		
KM	Vaccine: eye disease; conventional trachoma; nonendemic trachoma;		
KM	paratrachoma; inclusion conjunctivitis; genital disease; periphthalis;		
KM	nongonococcal urethritis; epididymitis; cervicitis; salpingitis;		
KM	Bartonellosis; pneumonia; venereal lymphogranulomatosis.		
OS	Chlamydia trachomatis.		
PV	MO9928475-A2.		
ED	10-JUN-1999.		
FJ	27-NOV-1998;	98WO-IB01939.	
XX	04-NOV-1998;	98US-0107077.	
PR	28-NOV-1997;	97ER-0015041.	
PR	17-DEC-1997;	97FR-0016034.	
PA	(GENSET ) GENSET.		
PI	Griffais R,		
DR	WPI; 1999-371125/31.		
XX	Genome sequence of Chlamydia trachomatis		
PS	Disclosure; Page 836; 1755pp; English.		
CC	AAV6754-Y37949 are encoded by open reading frames (ORFs) of the genome		
CC	of Chlamydia trachomatis (see AAZ01425). The polypeptides can be used as		
CC	vaccines against Chlamydia trachomatis. Antigenic and ribozyme sequences		
CC	can also be used to control growth of the microorganism. Chlamydia		
CC	trachomatis is responsible for a large number of sexually transmitted		





XX W09928475-A2.  
 PN  
 XX 10-JUN-1999.  
 PD  
 XX  
 XX 27-NOV-1998; 98WC-IB01939.  
 PF  
 XX  
 PR 04-NOV-1998; 98US-0107077.  
 PR 28-NOV-1997; 97FR-0015041.  
 PR 17-DEC-1997; 97FR-0016034.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Griffiths R;  
 XX  
 DR WPI: 1999-371125/31.  
 XX  
 PT Genome sequence of Chlamydia trachomatis  
 PS  
 XX Disclosure; Page 853; 1755pp; English.  
 XX  
 CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
 CC of Chlamydia trachomatis (see AAY201425). The polypeptides can be used as  
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
 CC can also be used to control growth of the microorganism. Chlamydia  
 CC trachomatis is responsible for a large number of diseases, e.g. eye  
 CC diseases such as conventional trachoma, nongonococcal trachoma,  
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
 CC and venereal lymphogranulomatosis. The polypeptides of the invention  
 CC may be of use in treating these diseases.  
 CC  
 XX  
 SQ Sequence 529 AA;

Query Match 24.8%; Score 680.5; DB 20; Length 529;  
 Best Local Similarity 34.0%; Pred. No. 5.8e-53;  
 Matches 179; Conservative 92; Mismatches 226; Indels 29; Gaps 10;

QY 14 LSLSVLQCKESSHSTSRGELAINIRDEPRSLDPRVRLSLVKTVEGLVQENN 73  
 DB 14 vslaiglsicynhqe--pkdvrliaichdpmldprvflskdvslvayeglvreke 71  
 QY 74 LSGNIEPLADYLSDDGLTFYFKLSAFNSGNDPLTAEDFESKMOVAQTOEGSIATF 133  
 DB 72 aa--fglaaerysqddgcyyffiknfwngdvrtaydfeeslkqlyfeidnpslr 129  
 QY 134 ALNPTKNRKIQDEHLSIDHFGVHSPNESTLVLTESPTSHFKLALPVPFVHKQSQT 193  
 DB 130 stailknshavltlgalpveldyralnaksleivlenpfyellelahpvtfstire 189  
 QY 194 L-----GSKSLP--TASGAFYPMKIKOKMIKSKNPHYVQSOVETKTTTHFTPDANTA 247  
 DB 190 ykkdkrnkrvfmnsnfpafiqcyepqrylllnknpiyhakhhvllnsvclqivpdihta 245  
 QY 248 AILFNQCKLNQGPWGERIPQETLSNLQSKGHSFDVAGTSMLEFNINRKNPLNMKLR 307  
 DB 250 mqlfqnkhidlvglpwssfsfleeqrnl-prekldfpvtsosvltcnlhqplnpslr 308  
 QY 308 EILASALDKALVSTIFLGRAKTADHLLPTNHSYPRHOKQEOAQAQAKLKEALEE 367  
 DB 309 talstalnretllk--laqkgcsatsfvhpqlsqipatlslq--derialakgyllealkt 365  
 QY 368 LQITAKDLEHLNIFPVSSSSASLLVOLIPEQKESGFAIPYVGERKALQADLSSGNF 427  
 DB 366 l--sgedlekltllyiesvcliravgeitqqlfdvqifgiklslqleyhcfldktrsgel 423  
 QY 428 SLATGGMFADFAPNAFLITAFPSGVPYPAIHNKDFLET-----LQNIQEDDQHKRS 461  
 DB 424 slatgmwladqbasafisvlg-----nglrykdfqldngnqkynhvaqillqess 476  
 QY 482 EIVSQASVLEFTHIIEPIYHDAFOFAMKKNLSNLGVSPGVVDER 527

DB 477 dlqmaeqllikespliplyhldvyvakqrvsdlqtsisgeidlk 522

## RESULT 7

AAY97040  
 ID AAY97040 standard; Protein; 544 AA.  
 XX  
 AC AAY97040;  
 XX

31-OCT-2000 (first entry)

B. subtilis oppA ligand binding protein.

Opp operon: Spook; oligopeptide permease; sporulation; ABC transporter;  
 ATP-binding cassette transporter; mutation; protein production; oppA;  
 oppB; oppC; oppD; oppE; ligand binding protein.

Bacillus subtilis.

W0200039323-A2.

06-JUL-2000.

21-DEC-1999; 99WC-US31010.

24-DEC-1998; 98GB-0028711.

(GENM ) GENENCOR INT INC.  
 (DIAZ/) DIAZ-TORRES M.  
 (FERR/) FERRARI E.

Diaz-Torres M, Ferrari E;  
 WPI: 2000-452412/39.  
 DR  
 N-PSDB: AAA51864.

Production of proteins, such as hormones, enzymes, growth factors or  
 cytokines, in gram-positive microorganisms containing a mutation in at  
 least one of the genes of the opp operon gene cluster

Disclosure; Fig 1A-W; 32pp; English.

The opp operon of Bacillus (also known as spook operon) encodes an  
 oligopeptide permease that is required for the initiation of sporulation  
 and the development of genetic competence. The opp operon is a member of  
 the family of ATP-binding cassette (ABC) transporters involved in the  
 import or export of oligopeptides from 3-5 amino acids. Bacillus strains  
 containing a mutation in the opp operon produce more recombinant protein  
 than the wild-type strain. Therefore, gram-positive microorganisms,  
 especially Bacillus strains, containing a mutation in at least one of the  
 genes of the opp operon can be used for heterologous protein production,  
 especially hormones, enzymes (preferably proteases such as subtilisin),  
 growth factors or cytokines.

Sequence 544 AA;

Query Match 22.5%; Score 617.5; DB 21; Length 544;  
 Best Local Similarity 29.4%; Pred. No. 3.3e-47;  
 Matches 159; Conservative 107; Mismatches 249; Indels 25; Gaps 13;

QY 10 ITLLSLVVLQCKESSHSTSRGE-----LAINIRDEPRSLDPRVRLSLV 61  
 DB 8 vtlmlitflvlsaagfygssnggskakgtctlnlnkicpftslpplandsvsgvyl 67  
 QY 62 KHIVGIVQENNLSGNIEPALADYLSDDGLTFYFKLSAF--WSNGDPLTAEDFESGMK 120  
 DB 68 rglfegltlr--lnadgpeegmskietakdktlyftlrdgvkwsngdpytqgdeyawk 126  
 QY 121 QVAIEVSGIYAFALNPIKWKRIQDEHLSIDHFGVHSPNESTLVLTESPTSHFKLTA 180  
 DB 127 walpnmessgyayqlylykgaeanantkgyslddvavkavndclikvelnnpfyftelta 186

RESULT	8
AA37004	
ID	AA37004 standard; Protein: 214 AA.
AC	AA37004;
XX	
DT	07-OCT-1999 (first entry)
XX	
DE	Chlamydia trachomatis transport protein.
XX	
KM	Vaccine; eye disease; conventional trachoma; nongonococcal urethritis; epididymitis; genital disease; perinephritis; Bartholinitis; pneumonia; venereal lymphogranulomatosis.
KM	paratrachoma; inclusion conjunctivitis; genital disease; perinephritis; Bartholinitis; pneumonia; venereal lymphogranulomatosis.
KM	Bartholinitis; pneumonia; venereal lymphogranulomatosis.
XX	
OS	Chlamydia trachomatis.
XX	
PN	W0928475-A2.
XX	
PD	10-JUN-1999.
XX	
PF	27-NOV-1998; 98WO-IB01939.
XX	
PR	04-NOV-1998; 98US-0107077.
PR	28-NOV-1997; 97FR-0015041.
PR	17-DEC-1997; 97FR-0016034.
XX	
PA	(GEST) GENSET.
XX	
PI	Griffais R.
XX	
DR	WPI; 1999-371125/31.
XX	
PT	
XX	
PS	Genome sequence of Chlamydia trachomatis
XX	
PS	Disclosure: Page 836; 1755pp; English.
XX	
AY36754	AY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
CC	of Chlamydia trachomatis (see A201425). The polypeptides can be used as
CC	vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC	can also be used to control growth of the microorganism. Chlamydia
CC	trachomatis is responsible for a large number of diseases, e.g. eye
CC	diseases such as conventional trachoma, nongonococcal urethritis,
CC	paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC	nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
CC	perinephritis, Bartholinitis; pneumonia; venereal lymphogranulomatosis.
CC	The polypeptides of the invention
CC	and venereal lymphogranulomatosis. The polypeptides of the invention

```
Query Match      19.0%; Score 521.5; DB 20; Length 214;
Best Local Similarity 48.5%; Pred. No. 4.2e-39;
Matches 98; Conservative 41; Mismatches 62; Indels 1; Gaps 1;
```

RESULT	9
AAY20067	
ID	AAY20067 standard; Protein; 512 AA
XX	

B. burgdorferi antigenic protein, t606.aa.

Antigenic protein; vaccine; Lyme disease; infection; detection.

OS *Borrelia burgdorferi*  
XX

PN W09859071-A1.  
XY

PD 30-DEC-1998.

18-JUN-1998; 98WO-US12718

XX 03-SEP-1997. 02115-0057403  
PR

PR 20-JUN-1997; 97US-0050359  
 PR 20-JUN-1997; 97US-0050359

22-JUL-1997; 97US-0053377

AA  
PA  
(HUMA-) HUMAN GENOME SCI INC

PA (MEDI-) MEDIMMUNE INC.  
XX

PI Choi GH, Erwin AL, Hanson MS, Lathigra R;  
XX

DR WPI; 1999-189980/16.

DN N-PSDB; AAX01/64.  
XX

PT New isolated Borrelia products for the dia

caused by *Borrelia*,  
PT  
XX

PS Claim 12; Page 182;  
xx

CC This sequence repres

cc invention, which is  
cc can be used in vacui

CC the Borrelia genus,

infection caused by

cc be used for detection  
xx

This sequence represents a *Borrelia burgdorferi* (Bb) protein of the invention, which is suitable for use in a vaccine. The Bb polypeptides can be used in vaccines for eliciting protective antibodies to members of the *Borrelia* genus, particularly for the use against Lyme disease in humans and animals. They can be used for preventing or attenuating an infection caused by a member of the *Borrelia* genus. The products can also be used for detection of members of the *Borrelia* genus.

100



Db 120 ftywskklvlpitgnaylldsvknsfelrngeksvdelglsapndkefiwelkqags 179  
 QY 175 FLKLALPVPFVHKSORLQSKSLP-----IASGAFYFKN--IKOKWIKLSKNPHYY 226  
 Db 180 flavvslawlapngkfveagqkdyaldsehllyspgfllanwdatsdtw-tilkknpeyy 238  
 QY 227 NOSQVERKTTIFIPIDANTAKLFLNQGKLMWQCPFWGGERIPQETLSNQ-SKGHLHSFD 265  
 Db 239 dadqvkleeavastikedtgnilyvneidl-----vrlngqyvqgqddppgyvshpd 292  
 QY 286 VAGTSWLTFINIK-FPLNNMKLREALASALDKEALVSTIFLGRATADHLPTNHSYRE 344  
 Db 293 va-nylfdnkkkeqrpianlvhrkaigaldkeelqsvlndgspnrlngllipsklyanpe 351  
 QY 345 HOKEMAQRGAYAKKILFEALBELQITAKDL-EHLNL-IFPVSSASSLLVQLIREQWKE 402  
 Db 352 tdefefraysgelykndvkkagaewtkagadvgrkvlslaaadtqgkxiieyqsglqe 411  
 QY 403 SL-GFAIPYVKGFEFALLQADLSGNGFSLATGWFADFPAPMAFLTIFAYPGVPPYAINH 461  
 Db 412 nlpglellssqpsnmvngstreknyelstsgwlaqsseldeslylnlygeessy-nygnyh 470  
 QY 462 K---DFLEILONIEQDQHOKRSELVSOASLYLFTFHIIEPIYHDAROFAMNKKLSNL 516  
 Db 471 nakydqiveearllnanmpkqfaeykkaedlllngdaaqvplygsasnylinpklygi 529

## RESULT 13

AAV00031 standard; Protein; 528 AA.

AAV00031;

20-APR-1999 (first entry)

Enterococcus faecalis antigenic polypeptide fragment EF012.

Enterococcus faecalis; infection; vaccine; immune response; diagnosis;  
 detection; attenuation; antigenic.

Enterococcus faecalis.

W09850554-A2.

12-NOV-1998.

04-MAY-1998; 98WO-US08959.

14-NOV-1997; 97US-0066009.

06-MAY-1997; 97US-0044031.

16-MAY-1997; 97US-0046655.

(HUMA-) HUMAN GENOME SCI INC.

Bailey C, Choi GH, Hromockyj A, Kunsch CA;

WPI; 1999-070095/06.

DR N-PSDB; AAX20021.

New isolated Enterococcus faecalis polynucleotides - used to develop  
 products for the detection of Enterococcus and for use in vaccines  
 for prevention or attenuation of Enterococcus infection

Claim 9; Page 91; 301pp; English.

The present sequence represents an antigenic polypeptide fragment  
 isolated from Enterococcus faecalis. The present invention describes  
 genes, proteins and antigenic polypeptides isolated from E. faecalis.  
 The proteins can be used in vaccines for preventing or attenuating an  
 infection caused by a member of the Enterococcus genus in an animal.  
 They can also be used for detecting Enterococcus antibodies in a sample.  
 The nucleotide sequences can be used for detecting Enterococcus nucleic  
 acids. Products from the present invention can also be used for

CC screening compounds to identify agonists and antagonists of E. faecalis  
 CC protein activity.  
 XX  
 SQ Sequence 528 AA;

Query Match 16.5%; Score 452.5; DB 20; Length 528;  
 Best local similarity 26.6%; Pred. No. 3,36-32;  
 Matches 138; Conservative 102; Mismatches 244; Indels 35; Gaps 17;

QY 22 GKESHSSTSRGELA---INIRDEP--RSLDPQVRLSEISLVHYEGHVGNNLS 75  
 Db 3 glkeaaekvds-gnlaaeqkisspspistldtqtdkntfmeaghlfeqlyrfddds 61  
 QY 76 GNIEPALAEYSSISDGLTYTFKLKSAF-WSNGDEPLAEDFISWQOVA TOEVSQYAF 134  
 Db 62 atv-palakdkkissddgklyhflfeqllkwsngeplaqdfvyswkklytpitignayl 120  
 QY 135 LNPILKNVAKIDEGHSLTDHGVHSPNESTLVVTLSEFSLKLALLPVPFVHKSORTL 194  
 Db 121 ldsvknsfelrngeksvdelgisapndkefiwelkqagsflavvslawlapngkfvea 180  
 QY 195 QSKSLP-----IASGAFYFKN--IKOKWIKLSKNPHYYNOSQVETKTTIHFIPOANT 246  
 Db 181 qkdyaidselhlysgpfllanwdatsdtw-tilkknpeyydaaqvavleavastikedt 239  
 QY 247 AAKLFNQGKLMWQCPFWGGERIPQETLSNQ-SKGHLHSFVAGTSMWLFNINM 304  
 Db 240 glnlyvneidl-----vrlngqyvqgqddppgyvshpdva-nylfdnkkkeqrpianv 292  
 QY 305 KLREALASALDKEALVSTIFLGRATADHLPTNHSYREHOKEMAQRGAYAKKILFEA 364  
 Db 293 hlraiaigaldkealqsvlndgskpnlngllipsklyanpedefraysgelykndvkk 352  
 QY 365 LEELOITAKDL-EHLNL-IFPVSSASSLLVQLIREQWKE-SGFAIPYVKGFEFALLQAD 421  
 Db 353 qawtkagadvgrkvlslaaadtqgkxiieyqsglqenlpglellssqpsnmngs 412  
 QY 422 LSSGNSLATGWFADFPAPMAFLTIFAYPGVPPYAINHK---DFLEILQNTQDQDH 477  
 Db 413 rreknyelstsgwlaqsseldeslylnlygeessy-nygnyhnaikyqjiveearllnanpe 471  
 QY 478 QKRSELVSOASLYLFTFHIIEPIYHDAROFAMNKKLSNL 516  
 Db 472 kqfaeykkaedlllngdaaqvplygsasnylinpklygi 510

## RESULT 14

AAV19870 standard; Protein; 523 AA.

AAV19870;

19-JUL-1999 (first entry)

B. burgdorferi antigenic protein, f607.aa.

Antigenic protein; vaccine; Lyme disease; infection; detection.

Borrelia burgdorferi.

W09859071-A1.

30-DEC-1998.

18-JUN-1998; 98WO-US12718.

03-SEP-1997; 97US-0057483.

20-JUN-1997; 97US-0050359.

22-JUL-1997; 97US-0053344.

22-JUL-1997; 97US-0053377.

(HUMA-) HUMAN GENOME SCI INC.



PA (MEDI-) MEDIMUNE INC.

PI Choi GH, Erwin AL, Hanson MS, Lathigra R;

DR WPI: 1999-189980/16.

DR N-PSDB: AAX61567.

PT New isolated *Borrelia burgdorferi* nucleic acids - used to develop  
PT products for the diagnosis, prevention and treatment of diseases  
caused by *Borrelia*, particularly Lyme disease

PS Claim 12; Page 97-98; 275pp; English.

CC This sequence represents a *Borrelia burgdorferi* (Bb) protein of the  
CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
CC can be used in vaccines for eliciting protective antibodies to members of  
CC the *Borrelia* genus, particularly for the use against Lyme disease in  
CC humans and animals. They can be used for preventing or attenuating an  
CC infection caused by a member of the *Borrelia* genus. The products can also  
CC be used for detection of members of the *Borrelia* genus.

XX Sequence 523 AA;

Query Match 15.5%; Score 425.5; DB 20; Length 523;

Best Local Similarity 23.4%; Pred. No. 9.2e-30;

Matches 122; Conservative 131; Mismatches 217; Indels 51; Gaps 15;

DB 1 MRKISGICITLLSVLVOGCKESHSSSTSGELAINIRDEPRSDPRVRLLEISL 60

DB 1 mkykialmliflsiacl-----snakektivrsnlsespsldpqlstldlygsnl 53

DB 61 VKHIVGLVOENNLNGNTEPALADYSLSDGILTYFKLK-SAEWSNDPLTAEPTESM 119

DB 54 ltlflglavkdsqgkykpglakswnlsedgilyfnlredlwsdgvateelkky 113

DB 120 KOVATQEVSGIYAFAL-NPIKNVAKIOEGHLSIDHGVHSPNESTIVVTEPTSHFLKL 178

DB 114 lrlnktaamyanaikstiknaeyfdetvpeselgikalskltltsltpkpyfpm 173

DB 179 LALVPEPV--HKSQRTLOSKSLP---IASGAFYPRKNKOKOWIKLSKNPHYNOQVET 233

DB 174 lthsaylpypmilvekygemwtnpenlivsgaykikersindkivleknkyyanaknvel 233

DB 234 KTTIHFIPIPDANTAKLFPNOGKLNW--QGPWGERIPQETLSNLOSKGHLSPFVAGTSL 292

DB 234 dev-ityptegsvaymyingeldlfg-----aeknleeikirdyysgikngmayl 286

DB 293 TFINKFPPLNMKTLREALASALDKREALVSTIFLGRAKTADHLPT-NHISY-----P 343

DB 287 afnttkpdlndkvrqaislaidrecltkvvlkgsdptnltpktdyysgknllldp 346

DB 344 EMOQDMARQAYAKKLFKALBELDQITAKDLEHNLFFPYSSASSLVLQULIREQKES 403

DB 347 enakkllaeagypdgkfp-----tlkykiseqrpt-taeflgeqfkk1 389

DB 404 LGFAPIYGRFALLQADLSSGFSLATGCGFADPAPMAFL-TIPAYPSG-VPPYAINH 461

DB 390 llnleeneewtflfgsrrtgnymssvgydgfdplfllslfthenhflgkyksn 449

DB 462 KDFLEILDINIEQODHOKRELYSOA-SLYETFHIIPEIY 501

DB 450 keydaliksnfeldpikrqiqlrqeeelaeakdfpmayl 490

RESULT 15

AAI19871

AAI19871 standard; Protein; 506 AA.

AC AAI19871;

DT 19-JUL-1999 (first entry)

XX

DE B. burgdorferi antigenic protein, t607.aa.

KM Antigenic protein; vaccine; Lyme disease; infection; detection.

OS *Borrelia burgdorferi*.

XX W09859071-AL.

XX 30-DEC-1998.

XX 18-JUN-1998; 98MO-US12718.

XX 03-SEP-1997; 97US-0057483.

XX 20-JUN-1997; 97US-0050359.

XX 22-JUL-1997; 97US-0053344.

XX 22-JUL-1997; 97US-0053377.

XX (HUMA-) HUMAN GENOME SCI INC.

PA (MEDI-) MEDIMUNE INC.

PA Choi GH, Erwin AL, Hanson MS, Lathigra R;

DR WPI: 1999-189980/16.

DR N-PSDB: AAX61568.

PT New isolated *Borrelia burgdorferi* nucleic acids - used to develop  
PT products for the diagnosis, prevention and treatment of diseases  
caused by *Borrelia*, particularly Lyme disease

PS Claim 12; Page 98; 275pp; English.

CC This sequence represents a *Borrelia burgdorferi* (Bb) protein of the  
CC invention, which is suitable for use in a vaccine. The Bb polypeptides  
CC can be used in vaccines for eliciting protective antibodies to members of  
CC the *Borrelia* genus, particularly for the use against Lyme disease in  
CC humans and animals. They can be used for preventing or attenuating an  
CC infection caused by a member of the *Borrelia* genus. The products can also  
CC be used for detection of members of the *Borrelia* genus.

XX Sequence 506 AA;

Query Match 15.4%; Score 421; DB 20; Length 506;

Best Local Similarity 24.6%; Pred. No. 2.2e-29;

Matches 118; Conservative 119; Mismatches 198; Indels 44; Gaps 14;

DB 43 EPRSLDPROVRLISRLSVKHYGLVOENNLNTEPALADYSLSDGILTYFKLK-S 101

DB 19 epsldpqlstldlygsnlitnlflglavkdsqgkykpglakswnlsedgilyfnlred 78

DB 102 AFMSNDPLTADPTESKQVATQEVSGIYAFAL-NPIKNVAKIOEGHLSIDHGVHSPN 160

DB 79 iwsdgvataeeikysylrlnktaamyanaikstiknaeyfdetvpeselgikald 138

DB 161 ESTLVLTESPTSHFLKLLALPVPEPV--HKSQRTLOSKSLP---IASGAFYPRKNKOKO 215

DB 139 skleltltspprypdmllthsaylpypmilvekygemwtnpenlivsgaykikersind 198

DB 216 WIKLSKNPHYNOQVETKTTIHFIPIPDANTAKLFPNOGKLNW--QGPWGERIPQETLSN 274

DB 199 klvleknkyyanaknveldev-ityptegsvaymyingeldlfg-----aeknlee 251

DB 275 LOSGHLSPFVAGTSMWLNKFPPLNMKTLREALASALDKREALVSTIFLGRAKTADHL 334

DB 252 ikirdyysgikngmayiafntlkpdlndkvrqaislaidrecltkvvlkgsdptnl 311

DB 335 LPT-NHISY-----PEHOKQDMARQAYAKKLFKALBELDQITAKDLEHNLFFPY 385

DB 312 tpfktdyysgknlllfpdenakkliaagypdgkfp-----tlkykis 355

DB 386 SSASSLVLQULIREQKESLGFAPIYGRFALLQADLSSGFSLATGCGFADPAPMAFL 445

DB 356 egrpt-taeflgeqfkkllnleeneewtflfgsrrtgnymssvgydgfdplfll 414

Fri Jul 26 08:37:37 2002

us-09-824-567-2.rag

Page 11

QY 446 -TIFAYPSG-VPEYAINHKDFLEITQNEQODHQRSELSOA-SLYLETFHIIIEPIY 501  
Db 415 dslfthenhlgaykysnkeydaliksnfelbpbikrtdllrqaeelisekdmpmaly 473

Search completed: July 26, 2002, 04:38:09  
Job time: 5787 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2002, 03:34:32 ; Search time 32.07 Seconds  
(Without alignments)  
405,189 Million cell updates/sec

Title: US-09-824-567-2

Perfect score: 2739

Sequence: 1 MRISVIGICITILSLSVL.....LSNMGVSPGVDFRYAKEN 532

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents\_AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCYUS.COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match length	DB ID	Description
1	280.5	10.2	642	US-08-245-511-48
2	280.5	10.2	642	US-08-600-993A-48
3	270.5	9.9	643	US-08-245-511-47
4	270.5	9.9	643	US-08-600-993A-47
5	165.5	6.0	556	US-09-385-028-9
6	108	3.9	398	US-08-935-263-12
7	107	3.9	1089	US-08-375-300-2
8	107	3.9	1089	US-09-177-431-2
9	107	3.9	1089	US-08-328-256-10
10	104.5	3.8	557	US-08-471-454-2
11	104.5	3.8	557	US-08-466-974-2
12	104.5	3.8	557	US-08-471-453-2
13	104.5	3.8	557	US-08-307-588-4
14	104.5	3.8	557	US-09-415-522-6
15	104.5	3.8	2071	US-08-354-973-1
16	103	3.7	436	US-08-660-347-2
17	103	3.7	2237	US-08-354-973-1
18	100.5	3.7	580	US-08-049-783-2
19	100	3.7	1257	US-08-158-232-6
20	100	3.7	1257	US-08-304-626-6
21	100	3.7	1257	US-08-316-301A-6
22	100	3.7	1257	US-08-611-928-6
23	100	3.7	1257	US-09-173-891-6
24	100	3.7	1257	US-09-076-137-6
25	100	3.7	1257	PCT-US92-03624-6
26	100	3.7	1257	US-08-378-313-34
27	99	3.6	476	Sequence 48, Appl

28	97	3.5	1664	US-08-642-846-2	Sequence 2, Appl
29	97	3.5	1664	US-09-264-604-2	Sequence 2, Appl
30	97	3.5	2710	US-08-480-604A-6	Sequence 6, Appl
31	97	3.5	2710	US-08-405-496A-6	Sequence 6, Appl
32	97	3.5	2710	US-08-915-136-6	Sequence 6, Appl
33	94	3.4	1142	US-08-993-118-7	Sequence 7, Appl
34	94	3.4	1142	US-08-845-528C-7	Sequence 7, Appl
35	94	3.4	1142	US-09-061-709-2	Sequence 7, Appl
36	93.5	3.4	2629	US-08-751-189-4	Sequence 2, Appl
37	93.5	3.4	2629	US-09-060-836-4	Sequence 4, Appl
38	93.5	3.4	2629	US-09-184-445-4	Sequence 4, Appl
39	92.5	3.4	893	US-07-977-434-4	Sequence 4, Appl
40	92.5	3.4	893	US-08-458-819-4	Sequence 4, Appl
41	92.5	3.4	893	US-09-105-697-10	Sequence 4, Appl
42	92.5	3.4	893	PCT-US91-07035-4	Sequence 10, Appl
43	92	3.4	512	US-08-856-253-6	Sequence 6, Appl
44	91.5	3.3	920	US-08-930-996A-8	Sequence 8, Appl
45	90	3.3	776	US-08-961-083-86	Sequence 86, Appl

## ALIGNMENTS

RESULT 1  
US-08-245-511-48  
Sequence 48, Application US/08245511  
Patent No. 5928900  
GENERAL INFORMATION:  
APPLICANT: Masure, H Robert  
APPLICANT: Pearce, Barbara J  
APPLICANT: Tuomanen, Elaine  
TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND  
TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON  
NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Klauder & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/245,511  
FILING DATE: 18-MAY-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/116,541  
FILING DATE: 01-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-069 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 48:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 642 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
IMMEDIATE SOURCE:  
CLONE: am1A

```

FEATURE:
OTHER INFORMATION: NOTE: the reference contains a
OTHER INFORMATION: sequence error; the correct sequence shown below is obtained
OTHER INFORMATION: from GENBANK
PUBLICATION INFORMATION:
AUTHORS: Allouing, et al.
JOURNAL: Mol. Microbiol.
VOLUME: 4
PAGES: 633-644
DATE: 1990
US-08-245-511-48

Query Match      10.2%; Score 280.5; DB 2; Length 642;
Best Local Similarity 21.7%; Pred. No. 6,1e-18;
Matches 137; Conservative 104; Mismatches 243; Indels 147; Gaps 26;

QY 19 VLQCKESHSTSGELAIN--IRDEPRSLDPQVRLISELSVYKHIEGLVQENNLISG 76
D 2 VLAC--SSKSSSDSAPKAGVYTTADPETIDYLSRKNSVTWVTSNGIDGLFTNDWY-G 59
QY 77 NIEPALADYSLSDGLTYTFKLSA--FWSNGD--PLTAEDFIESMKOVATQEVSGI 130
D 60 NLAPAAVEDMEVSKDGLTYTKIRKGVKMTSDGEYAETAKDVYNGIKHAADKKSEAM 119
QY 131 YAFALNPKNVKTQEG-HLSIDHGVHSPNESTLVWLESPTSHFLKLALPVFFPVK 189
D 120 Y-LAENSVKGLADYLSGTSTDFSVGVKAVDDYTLQTLNQPPEPNWSKLTVSIFWPLNE 178
QY 190 SORTLOSKSLP-----IASGAFYPRKNIKQKWKLSKNPHYVNSQVETKTTIT--- 238
D 179 EETSCKSDFAKPTDPTSLYNKPFLLKGLTRKSSVEPVKNEQYWDKEVHLDITNLAYY 238
QY 239 -----HETPDANTAKLFNOGKLNMOGPPWGBRIPQETLSMLQSKGHLHSDVAG 288
D 239 DGSODESLERWFTSGAVYARLYPTSS-NY-----SKVAEYKNDI-----YTIQSGSG 286
QY 289 TSMLEFNINKEPLN-----NKKLEALASLDKEA----- 318
D 287 IAGLGVNIDQSYNTYTSKTTTDESEKVAKTKALLNDFRQALNFDLSRAYSQAQINGKDGAA 346
QY 319 -LVSTIPL-----GAKTADHLPTLNHISYPEHQ-----QEAQAQAATAKTLKEA 364
D 347 LAVRLEVPKPPFVSAKEKTFEGDLVAAQPAYGDEMKKVNADQODLPNADKAAEFKKA 406
QY 365 LEELOITAKMLEHLNIFPVSSASSLL--VOLIRQWKESSGFAIPVY-----GKEF- 415
D 407 KKALBADGVOP-THLDVAVDASKNYISRTQSFQSEVETVGVENVVYDQWMTDEFL 465
QY 416 --ALLQADLSGNSFLATG-GWAFADFPAPMAFLTI-----PAYPGVPPYA 458
D 466 NITYYANMASSEMDVSSGVSGVSPYODPTSTYLDILKTTYSSETTYTGFPNPNPSVYQ 525
QY 459 INHNDPELLONTQDE-ODHOKRESELYSOASLYLETFHIEPIYHDAFQPMNKKSLNIG 517
D 526 VGLKEYDKLVDAAKETSDFNVRKEXYAAQAAMLTDSLSLFI-----AASGSG 573
QY 518 VSP-----TGVS-----YDPRYAK 530
D 514 AAPVLSRTVPTGASQGTGSGSDVYFKYK 604

RESULT 2
US-08-600-993A-48
; Sequence 48; Application US/08600993A
; Patent No. 5981229
; GENERAL INFORMATION:
; APPLICANT: Masure, H Robert
; APPLICANT: Pearce, Barbara J
; APPLICANT: Tuomanen, Elaine
; TITLE OF INVENTION: BACTERIAL EXPORDED PROTEINS AND
; TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON
; NUMBER OF SEQUENCES: 59

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/600,993A
FILING DATE: 1-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/245,511
FILING DATE: 18-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/116,541
FILING DATE: 01-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-069 US
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 642 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHEICAL: NO
AMTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
IMMEDIATE SOURCE:
CLONE: amla
FEATURE:
OTHER INFORMATION: the reference contains a sequence error; the
PUBLICATION INFORMATION:
AUTHORS: Allouing, et al.
JOURNAL: Mol. Microbiol.
VOLUME: 4
PAGES: 633-644
DATE: 1990
US-08-600-993A-48

Query Match      10.2%; Score 280.5; DB 2; Length 642;
Best Local Similarity 21.7%; Pred. No. 6,1e-18;
Matches 137; Conservative 104; Mismatches 243; Indels 147; Gaps 26;

QY 19 VLQCKESHSTSGELAIN--IRDEPRSLDPQVRLISELSVYKHIEGLVQENNLISG 76
D 2 VLAC--SSKSSSDSAPKAGVYTTADPETIDYLSRKNSVTWVTSNGIDGLFTNDWY-G 59
QY 77 NIEPALADYSLSDGLTYTFKLSA--FWSNGD--PLTAEDFIESMKOVATQEVSGI 130
D 60 NLAPAAVEDMEVSKDGLTYTKIRKGVKMTSDGEYAETAKDVYNGIKHAADKKSEAM 119
QY 131 YAFALNPKNVKTQEG-HLSIDHGVHSPNESTLVWLESPTSHFLKLALPVFFPVK 189
D 120 Y-LAENSVKGLADYLSGTSTDFSVGVKAVDDYTLQTLNQPPEPNWSKLTVSIFWPLNE 178
QY 190 SORTLOSKSLP-----IASGAFYPRKNIKQKWKLSKNPHYVNSQVETKTTIT--- 238

```

```

Db 179 EFETSKSGDPKPTDPTSLLYNGPFLKGLTAKSSVEFEKRNQYMKENVHLDITNLAY 238
QY 239 -----HPIPDANTAKLFPNGKILNMGPPWGERITQETLSLQSGHLSFDVAG 288
Db 239 DCSQDESIERNFTSGAYSTARLYPTSS-NY-----SKVAEEKQNI-----YTYQSGSG 286
QY 289 TSWLTFNINFPPLN-----NMKLEALASLQKEA-----318
Db 287 IAGGVNIDRQSYNTSKTTDEKATKALKLNKDFRQALNFRALDRSAVSQINGKDGAA 346
QY 319 -LVSTITL-----GRATADHLPTNHSYPERHOK-----QEMARQAVAKKLPREA 364
Db 347 LAVNLEFVKEDFVSAGEKTFGLVLAQAPAYGDEMKGVNLAQOGGILNNAKAAERKA 406
QY 365 LEELQITAKDLEHLNLEPYSSASLSL--VQILREQKESIGPAIPY-----GKEF- 415
Db 407 KKALEADGVOP-ITHLDVVDQASKNYISRSQSFQSVETVLGVENVVVDIQQMSDFL 465
QY 416 -ALLQADSSGNSFLATG-GMFADPADPMAFLTI-----FAYPSGVPPYA 458
Db 466 NITYAANASSEDMDVSGVSGWGPYQDPSTYIDLKTSSTTYTIGFNPSPSVVO 525
QY 459 INHDFELLLQNIQEG-ODHOKRSELYSQASLYLETFHIEPIYHDAFQFAMNKLNLG 517
Db 526 VGLKEIDKLVDAKETSDFNVRKYAAQAWLITDSLSLIP-----AASSG 573
QY 518 VSP-----TGV-----VDFRYAK 530
Db 574 AAPVLSRIVPTFGASAGTSGSGSVYTKYK 604

RESULT 3
; Sequence 47, Application US/08245511
; Patent No. 5928900
; GENERAL INFORMATION:
; APPLICANT: Masure, H Robert
; APPLICANT: Pearce, Barbara J
; APPLICANT: Tuomanen, Elaine
; TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND
; TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/245,511
; FILING DATE: 18-MAY-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/116,541
; FILING DATE: 01-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-069 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643 amino acids
; TYPE: amino acid

```

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-245-511-47

Query Match 9.9%; Score 270.5; DB 2; Length 643;  
Best Local Similarity 21.6%; Pred. No. 5,4e-17;  
Matches 130; Conservative 101; Mismatches 214; Indels 157; Gaps 25;

```

QY 19 VIQCKESHSTSRGELAINIDEPKSIDPROVLLSEISLVKHYEGIVQENNLGNI 78
Db 2 VLAACSGSGSSAGEKTFYSIYETDPDNLNLTAKATANTISNVVDGL-ENRIGNF 60
QY 79 EPALAEYSLSDGLTYTEKL-KSAFW--SNQD---PLTADPIESKQVAAQEVSGIYA 132
Db 61 VPSMAEDSVSKDILITTYTIIRKDAKWTISGEYAAKADQVFTGLKYAAKKSADALYP 120
QY 133 FALNDIKVNRIGQHL-SIDHGVHSPNESTIVLTLESPHSFKLLALPVEFPVHKSQ 191
Db 121 -VQESIKGLDAYVGEIKDFQVGIKALDEQYQYTLINKPEFSWNSKTTMGVLAV--NE 177
QY 192 RTLOSK-----SLPIASGATPKATIKOKOMIKLSKNPHYVQSOVERKTIITIFI 241
Db 178 EPLNSKGDPRKATIDSSSLYNGPYLLKSIYTKSSVEFAKNPNYMDKDHIDKYLSEFW 237
QY 242 --PDANTAALKLFGQKLNMGPPWGERI-----PQETL-----S 273
Db 238 DGQITSRFAENFKDGLT-----AARLYPSASFALSKSMKDNIVTQDSDITYLVGT 291
QY 274 NLQSKGHLHSFDVAGTSMLEFNKFPPLNNKLREALASALDEKALVSTI--FLGRAKTA 331
Db 292 NIDRQSYKTYT---SKTSDQKASTKALKLNKDFRQALAFQFDRTAAVQSLNQGKASKIL 348
QY 332 DHL--PTNHS-----YPERHQEMAQOQAY 356
Db 349 RNLVPTFYQADKNGKNGVKEKLVTVGDEMKDVNLADSDGLYNPEKAAERAKA--- 405
QY 357 AKLLEKALELEIQTAKDLEHLNLEPYSSASLSL--VOLLREQKESLG----- 405
Db 406 ---KSAQABEVTP---ITHDMPDQATATIKVORQOSKLEATLADAVIIDQ 456
QY 406 -----FALPIVKEFALLQADSSGNSFLATGGMFADPADPMAFL----- 445
Db 457 QLOKDEVNNTYFAERAAGEDM-----DLSD-----NNGMCPDADASTYIDLIRKSGV 505
QY 446 ---TIFAYPSG---VPPYALNHKDFLEIQLNT-EOGDHOKRSELYSQASLYLETFHII 497
Db 506 ESTKTYLGFDSGSDVNAKKVGLDYDEKLVTEAGDETTDVAAKRYKYAAQAWLITDSALI 565
QY 498 EP 499
Db 566 IP 567

RESULT 4
; Sequence 47, Application US/08600993A
; Patent No. 5981229
; GENERAL INFORMATION:
; APPLICANT: Masure, H Robert
; APPLICANT: Pearce, Barbara J
; APPLICANT: Tuomanen, Elaine
; TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND
; TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/600,993A
FILING DATE: 1-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/245,511
FILING DATE: 18-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/116,541
FILING DATE: 01-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-069 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 643 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-600-993A-47

Query Match          9.9%; Score 270.5; DB 2; Length 643;
Best Local Similarity 21.6%; Pred. No. 5.4e-17;
Matches 130; Conservative 101; Mismatches 214; Indels 157; Gaps 25;

19 VLQCKESHSTSRGELAINIRDEPRSLDPQVRLSPISLVKHYEGVQENLSGN 78
2 VLAACGGSSAKGKETSIEYEDPNDLNYLTAKAATANTSVNDGL-ENDRYGN 60
79 EPLADYSLSDGLITTEFL-KSAFY--SNGD--PLTAEDEFSMKOVAQEVSGIYA 132
61 VPSNADMSVSKDGLTYTIRKDAKWYSEGEVAAVQAQDFVGLKYAADKKSALY 120
133 FALNPLKVRKTOEGHL-SIDHFGVHSPNSTLYVLTSPSHFLKLLALVPFPPHKSQ 191
121 -VQESIKGIDAVYKGEIKDFQVGRALADQVYOTLTKPSPSFMNSKTTTGLVAPV--NE 177
192 RFLQSK-----SLPIASGAFYEPKNIKOKWIKLKNPHYNGSOYETKITITIFI 241
178 EFLNSKGDDEFAKATDPSLSLYNGPYLTKSTIVTKSSVEFAKNPNWQDNVHDKVLSFW 237
242 --PDANTAKLFGNCKLNMCGPPWGERI-----DOETL-----S 273
238 DGDYTSKPAENKRGDSL-----AARLYPTSASFALSKSMKDNIVYTQDSITYLVGT 291
274 NLSKGLHSPVACTSWLTPNINKPPLNMKLRALASALDKALVSTI--FLGRAKTA 331
292 NIDRQSYKT---SKTSDEOKASTKALLNKDFRQALAFGPDRTAVASQJLNGQTGSKIL 348
332 DILL--PTNHS-----YEHOKOMARQAY 356
349 RNLFPVPTFVADGKNFGDWVREKLYTYDEKKNVNLSDODGLYNPEKKALEFKA--- 405
357 AKTLREALEELQITAKDELHLIFPVSSASSLL--VOLIREQMKESIG----- 405
406 ----KSAIQEGVTFP---IHLDPVQDTATYKVOFQCMKQSLAEATIGADNVLIDIQ 456
406 ----FAPIVKEFALLQADSSCNFSLAGCFADFPADPAFL----- 445
457 QIQKDEVNNITTFAPENAGEDW-----DLSL-----NVGQDPDFADPSYLDIIPKSVG 505
446 ----TTFATPSG---VPPVAINHKDFLEILLONI-EQEQDHQRKSELVQASLYLETFPHLI 497

```

```

Db 506 ESTKTYLFDGCEGNVAKKVGLYDEKLYTAGDETTIDAKKIKVAAQAMLTDSALI 565
QY 498 EP 499
Db 566 IP 567

RESULT 5
US-09-385-028-9
Sequence 9, Application US/09385028
Patent No. 6232106
GENERAL INFORMATION:
APPLICANT: Susan E. Jensen
APPLICANT: Kwame A. Aidoo
APPLICANT: Ashish S. Paradkar
TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
Patent No. 6232106
TITLE OF INVENTION: Acid Biosynthesis
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC
STREET: The Jenner Building, 400 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/385,028
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/790,462
FILING DATE: 29-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: D. Douglas Price
REGISTRATION NUMBER: 24,514
REFERENCE/DOCKET NUMBER: 1418/P574520S2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-6666
TELEFAX: (202) 39305350
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 556 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-385-028-9

Query Match          6.0%; Score 165.5; DB 4; Length 556;
Best Local Similarity 21.0%; Pred. No. 4.2e-07;
Matches 97; Conservative 74; Mismatches 212; Indels 79; Gaps 17;

34 GELAINIRDEPRSLDPQVRLSPISLVKHYEGVQENLSG-----NIPALAEYSIS 89
29 GTLRIVRTDDEPSLDPGNITYYATWNLRLIGRTLYTPDTAPGKAGORLVPDLAESIGS 88
90 S-DGLTYTTRKLSAR-WSNGDPLTAEDEFSMKOVAQTOEVSGIYAFALNPLKVRKTOEG 147
89 SEDGRWYTRLEGRYEDGTPVVSADI---KHAJARSNGTDVLAGAP-----TYR 138
148 HLSIDHFG-----VHSPNSTLYVLTSPSHFLKLLALVPFPPHKSQRT- 193
139 HLLGTGCGPWRPEADQSPYLTLETIDERTIVRLRLNPPRAGMDLAIWSTTPVPRDRDYG 198

```



```

RESULT 6
US-08-935-263-12
: Sequence 12, Application US/08935263A
: Patent No. 6117669
: GENERAL INFORMATION:
: APPLICANT: FURUICHI, Yasuhiko
: APPLICANT: Hoshino, Tatsuo
: APPLICANT: Kimura, Hitoshi
: APPLICANT: Kiyasu, Tatsuya
: APPLICANT: Nagahashi, Yoshie
: TITLE OF INVENTION: BIOTIN BIOSYNTHETIC GENES
: FILE REFERENCE: Biotin Genes
: CURRENT APPLICATION NUMBER: US/08/935,263A
: CURRENT FILING DATE: 1997-09-22
: EARLIER APPLICATION NUMBER: EP 96115540.5
: EARLIER FILING DATE: 1996-09-27
: NUMBER OF SEQ ID NOS: 23
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 12
: LENGTH: 398
: TYPE: PR1
: ORGANISM: Kurthia sp.
US-08-935-263-12

```

Query Match	Best Local Similarity	3.9%;	Score 108;	DB 3;	Length 388;
Matches	82;	Conservative	50;	Mismatches 115;	Indels 122;
					Gaps 21;
QY 259	QGPWGERIPDET--LSNLSQKCHLSEPSVGTSMLEFNINKEPLNNMKLREPLASALD-	315			
Db	6 ILTPCEWKEIKKELAYLEIFESIQKRELVSYPFAQDWIM--INGKMLNLSANNYLQYAGDE	63			
QY 316	--KEALVSTIF-LIRAKTAPHLPLPIHSHSEPHOKDMA-----QROA-----	355			
Db	64 RLKKAADVAHTYAGATASRLI--IGNHPLQAOAQLVMNKKAEAGLIINSQYANL	120			
QY 356	-----YAKLKEKAL-----EELQTKADQLEHNLILEPVS--SASSL	392			
Db	121 GIISTLSRNDIYISDKLNHNASIVGALLSAKHLIRHNDLCHLPAALKKSSSEAKRL	180			
QY 393	V-----QLIR--EQW-----KESLQAFPIPVCKEFAIIQ	419			
Db	181 VIDTVPFSDGDFAYLEDTVLKERYENAMLTMDHAGSGIYKNGGEYA-----GHLHQQN	236			
QY 420	ADLSSGNFSLATGC-----WFADADPMALITFAPSGVPPYAIN-HKDFLEFIQ	469			
Db	237 IDIQGTSTKSLGSPGAYVVGKKMLIDLYKNR--MROFIYSTALDPALIGAMKATAEIYQ	294			
QY 470	NIEEQDHQKRSSELVSQASLYLET--FRI-----IEPTI-----HDAQOFAMKKIS	514			
Db	295 -----QEBEERSLQTHSEHFEEELTYVYGNGICSSQIPLVIVDENKAAHEFA--TRIQ	347			

```

RESULT 7
US-08-375-300-2
: Sequence 2, Application US/08375300
: Patent No. 5675566
:
: GENERAL INFORMATION:
:
: APPLICANT: Feng, He
:
: APPLICANT: Jacobson, Allan S.
:
: TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE PRODUCTION IN
:
: TITLE OF INVENTION: THE ABSENCE OF NONSENSE-MEDIATED mRNA DECAY FUNCTION
:
: NUMBER OF SEQUENCES: 6
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSES: Fish & Richardson
:
: STREET: 225 Franklin Street Suite 3100
:
: CITY: Boston

```

```

? ZIP: 02110-2804
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/375,300
? FILING DATE: 20-JAN-1995
? ATTORNEY/AGENT INFORMATION:
? NAME: Fasse, J. P.
? REGISTRATION NUMBER: 32,983
? REFERENCE/DOCKET NUMBER: 04020/046001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (617)542-5070
? TELEFAX: (617)542-8906
? TELEX: 200154
? INFORMATION FOR SEQ. ID NO.: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1089 amino acids
? TYPE: amino acid
? STRANDEDNESS: not relevant
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? US-08-375-300-2

```

```

Query Match: 3.98; Score 107; DB 1; Length 1089;
Best Local Similarity 18.68; Pred. No. 0.46;
Matches 112; Conservative 89; Mismatches 206; Indels 194; Gaps 30.

QY 52 VRLSLSTSLV-----KIIYGLVQENNLNGLNIEPALAEDYSLSDDLVTYFKL 99
DB 144 IRLVETLILVGYEFLRLDLIESKALPINEFLQKK--TGRRDPL--ESILRELINVKRGL 198
QY 100 ----KSAFWSGDPFLTAEDFIEFSPKQY-----ATQE-----VSGIYAPALNPILKY 141
DB 199 FTTTATATAIKKPAPFLPRD-DMSMDLLIYSKSLKGLQSLPFLNFIDATFAPADLHKRY 257
QY 142 RKIQEBSHLS-----IDHCVSHSPNESTLVLTLESPSHLKLIALPVFFPVH- 188
DB 258 NKLQRBHOKQIRGKLRLDYVEEY-----DKLLPIFIFKTS--AIIIGEFFMLEI 307
QY 189 ----KSORTLQSKSLPISAGAFYPPNIKOKW-----IKLSKMPHY 225
DB 308 PELEGASNDLDEKTASPMITNQLPPN--QLWMENEDTRKFTFILPDLISKIYESQSKT 365
QY 226 YNQSQVETKTTTHF-----TFD-----ANTAAKLFGNCKLIMQ 259
DB 366 EKDSVANSKINLFTDLLEMAQCKDIIDDLISNRYWSSYLDNKAIRRLRIKPFMTQ-DMS 424

```

QY 260 GPPMGR-----IPQETLSMLQS--KGLHSFDVAGTISWLTFFNINKPLNNKL 306  
 Db 425 KLPVYSRFLATNSKYMPEIVSEFINYLONGFRSOLHS-----NKINVKNT-- 469  
 QY 307 REALASALDKKALVSTYIFLG---RAKTAHLLPTNHSYF---EHOKQEMARQAYAKL 360  
 Db 470 --IFSEMIKQOLIPSMIFHKIRLIMQVNNVEILTVLHSGKFLNKPXY----- 523  
 QY 361 FREALAELOITAKDLEHLNLFPPVSSASSLLVOLIREQWESIGFAIPYGE-----FA 416  
 Db 524 -KELMEKVVQLIKDKKNDROLNNKMSALENIITLLKPPSKSLNTVATTIPEQGFRI 582  
 QY 417 LLOADLSSGNF---SLATGWFADPADMAFLTFAYP-----SGVPPY 457  
 Db 583 LRSELSLDEKHLVLYKVRKAWDVAIQVLFSLFSPKHSIQONIPILITVGLGLSY 642  
 QY 458 AINHKDFL-----EILONTF---QEDDHOKRSELVSAQSLYLFTH-----IIEPTY 501  
 Db 643 ---RDFVIRICDVLNLEIRGLEINDYGONNHRISVRYLTETEFEMIKSDVLDITLY 699  
 QY 502 H 502  
 Db 700 H 700

## RESULT 8

US-09-177-431-2  
 ; Sequence 2, Application US/09177431  
 ; Patent No. 6071700

## GENERAL INFORMATION:

APPLICANT: Jacobson, Allan S.  
 TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE PRODUCTION IN THE  
 NUMBER OF SEQUENCES: 18  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Richardson P.C.  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: Windows 95  
 SOFTWARE: PASCUSEQ for Windows Version 2.0b  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/177,431  
 FILING DATE:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/955,472  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Fasse, J. Peter  
 REGISTRATION NUMBER: 32,983  
 REFERENCE/DOCKET NUMBER: 07917/050001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617/542-5070  
 TELEFAX: 617/542-9806  
 TELEX: 200154  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1089 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-09-177-431-2

Query Match 3.9%; Score 107; DB 3; Length 1089;  
 Best Local Similarity 18.6%; Pred. No. 0.46;  
 Matches 112; Conservative 89; Mismatches 206; Indels 194; Gaps 30;

QY 52 VRLSEISLY-----KATYEGVOENNLSONIEPALAEDYLSDDGLTYTFKL 99  
 Db 144 LRVFTLYLVGFRTLDIESKDAIPNFIQKK--TGKDDPLL--FSILREIILNYKFLG 198  
 QY 100 -----KSAPWSNGDPLTAEDFIESMKQV-----ATQE-----VSGIYAFPLNIRKY 141  
 Db 199 FTTITATIKKFAFLPFRDD-DNSMDLLYDSKLGALQSLEKNNFIDATFARATELHKKV 257  
 QY 142 RKIOEGHLS-----IDHFGVSHNESTLVLTLESPTSHFLKLALVFPVPH- 188  
 Db 258 NKLOREHOKCOLRTGKLDEVEEY-----DKLPIPIRRTKS-----ATLLEEFKLEI 307  
 QY 189 ---KSOQTLOSKSLPIASATFPYNNIKQOM-----IKLSKNPHY 225  
 Db 308 PELEGASNDLKEHASPATITNOLIPN--QRLMENDTRKFEYIILPDISKTVESQSSKT 365  
 QY 226 YNOSQVETTYTIHF-----IPD-----ANTAALFMOGKLMQ 259  
 Db 366 EKDSNNVSKNINLPFTDLEMAQCKDIIDLSNRYSSYLONKATRNRIKFEWETQ-DMS 424  
 QY 260 GPPMGR-----IPQETLSMLQS--KGLHSFDVAGTISWLTFFNINKPLNNKL 306  
 Db 425 KLPVYSRFLATNSKYMPEIVSEFINYLONGFRSOLHS-----NKINVKNT-- 469  
 QY 307 REALASALDKKALVSTYIFLG---RAKTAHLLPTNHSYF---EHOKQEMARQAYAKL 360  
 Db 470 --IFSEMIKQOLIPSMIFHKIRLIMQVNNVEILTVLHSGKFLNKPXY----- 523  
 QY 361 FREALAELOITAKDLEHLNLFPPVSSASSLLVOLIREQWESIGFAIPYGE-----FA 416  
 Db 524 -KELMEKVVQLIKDKKNDROLNNKMSALENIITLLKPPSKSLNTVATTIPEQGFRI 582  
 QY 417 LLOADLSSGNF---SLATGWFADPADMAFLTFAYP-----SGVPPY 457  
 Db 583 LRSELSLDEKHLVLYKVRKAWDVAIQVLFSLFSPKHSIQONIPILITVGLGLSY 642  
 QY 458 AINHKDFL-----EILONTF---QEDDHOKRSELVSAQSLYLFTH-----IIEPTY 501  
 Db 643 ---RDFVIRICDVLNLEIRGLEINDYGONNHRISVRYLTETEFEMIKSDVLDITLY 699  
 QY 502 H 502  
 Db 700 H 700

## RESULT 9

PCT-US95-16930-2  
 ; Sequence 2, Application PC/TUS9516930  
 ; GENERAL INFORMATION:

APPLICANT: UNIVERSITY OF MASSACHUSETTS MEDICAL SCHOOL  
 TITLE OF INVENTION: HETEROLOGOUS POLYPEPTIDE  
 TITLE OF INVENTION: PRODUCTION IN THE ABSENCE OF  
 TITLE OF INVENTION: NONSENSE-MEDIATED MRNA DECAY  
 NUMBER OF SEQUENCES: 6  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Richardson  
 STREET: 225 Franklin Street Suite 3100  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US95/16930  
 FILING DATE: 27-DEC-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/375,300



QY 357 AKKLEALAELOITAKDLHLNLPVSSASSLLVLIQWKSGLGALPIVKEPA 416  
 Db 386 EKK-----TDVTPNLKPLT-VYCVARAHTM-----DEKLNKSVFSDAVCEK--- 428  
 QY 417 LLAADLSSGNFS--LATGGMFADFADPM-----AFLTIFAY--PSGVPPYALNH--- 461  
 Db 429 -----TKPGNTSKTWLIVGICIALFALPVIYAAKVFILRCINIVFPFSLKPSSTIDEXFS 483  
 QY 462 ----KDFL-----EILONT-----EEDODHOKRSELSVQAS 488  
 Db 484 EQLKNNLLSTSEOLEKCFIENISTLATVEETNGDDEHKKYSSQTSQDS 535

RESULT 11  
 US-08-471-454-2

Sequence 2, Application US/08471454  
 Patent No. 5731169

GENERAL INFORMATION:

APPLICANT: MOGENSEN, Knud E.

APPLICANT: UZE, Gilles

APPLICANT: LUTFALLA, Georges

APPLICANT: GRESSER, Ion

TITLE OF INVENTION: CDNA FRAGMENT CODING FOR THE GENE FOR

TITLE OF INVENTION: THE ALPHA INTERFERON RECEPTOR AND PROCESS FOR THE

TITLE OF INVENTION: PREPARATION OF THE CORRESPONDING PROTEIN

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHAYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/471,454

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/900,642

FILING DATE: 15-JUN-1992

APPLICATION NUMBER: FR 89/13770

FILING DATE: 20-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.

REGISTRATION NUMBER: 32,205

REFERENCE/DOCKET NUMBER: 960-7

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

TELEX: 200797 NIXN UR

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 557 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-471-454-2

Query Match 3.8%; Score 104.5; DB 1; Length 557;  
 Best Local Similarity 19.3%; Pred. No. 0.26;  
 Matches 114; Conservative 77; Mismatches 150; Indels 251; Gaps 33;

QY 18 VVLGCKSSSSSTSGELAINRDEPR-----SLDPPQV 52  
 Db 74 IKLSQNIISTKCNFSSILKNVYEIKIRIAKENNTSSWYVDSFTFPRRAQGPREV 133

QY 53 RLSE-ISLVKHYEGLVQENNLSGNTEPALADYSLSDGLTYTFLKASFNQDPLT 111  
 Db 134 HLAEBKALVTH-----ISPQKQSVAMALDGLSPYSL-LIWKNS---- 173  
 QY 112 AEDFISMQVATQVSGIYAFALNPKNRKIOEGHLSIDHFVSPNENSTLYVTFLESP 171  
 Db 174 -----SGVE-----ERENTYNSH-----KIYLSPE 195  
 QY 172 TSHFLKLA-----LPFPPVHKRSQPTLQSKSLPIASGAFYRNK----- 212  
 Db 196 TTYCLKVKFALLTSWTKIGVSPVHCITKYVENDELPP-----PENIEVSQNONVLKW 248  
 QY 213 -----QKWKI--LSKRP--HYNOSQVETKTIITHIPD--ANTAALFNQKRL 256  
 Db 249 DTYANMTFQVQWHAFLKRNPGNHLKWKQ-----IPDCENVKTKQCVFPQ--- 295  
 QY 257 NMGGPPWGRIRIPELTISLQSKG-----HLHSDVAGTSMWLFNPNK-----PLN 302  
 Db 296 -----NFKGITYLLRQASDGNNTSFWEEL-KDTETIQAFLLPVF 337  
 QY 303 NKLKEBALASLDKVALVSTFLGRAKTADHLPTNHSYP-----EHOKQEMARQAY 356  
 Db 338 NTR-----SLSDSFH-----IYIGAKQSGN-TVY-IDQYPLIYEIIFWENTSNMERKI 385  
 QY 357 AKKLEALAELOITAKDLHLNLPVSSASSLLVLIQWKSGLGALPIVKEPA 416  
 Db 386 EKK-----TDVTPNLKPLT-VYCVARAHTM-----DEKLNKSVFSDAVCEK--- 428  
 QY 417 LLAADLSSGNFS--LATGGMFADFADPM-----AFLTIFAY--PSGVPPYALNH--- 461  
 Db 429 -----TKPGNTSKTWLIVGICIALFALPVIYAAKVFILRCINIVFPFSLKPSSTIDEXFS 483  
 QY 462 ----KDFL-----EILONT-----EEDODHOKRSELSVQAS 488  
 Db 484 EQLKNNLLSTSEOLEKCFIENISTLATVEETNGDDEHKKYSSQTSQDS 535

RESULT 12

US-08-466-974-2

Sequence 2, Application US/08466974  
 Patent No. 5861258

GENERAL INFORMATION:

APPLICANT: MOGENSEN, Knud E.

APPLICANT: UZE, Gilles

APPLICANT: LUTFALLA, Georges

APPLICANT: GRESSER, Ion

TITLE OF INVENTION: CDNA FRAGMENT CODING FOR THE GENE FOR

TITLE OF INVENTION: THE ALPHA INTERFERON RECEPTOR AND PROCESS FOR THE

TITLE OF INVENTION: PREPARATION OF THE CORRESPONDING PROTEIN

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHAYE P.C.

STREET: 1100 NORTH GLEBE ROAD

CITY: ARLINGTON

STATE: VIRGINIA

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/466,974

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/900,642

FILING DATE: 15-JUN-1992

APPLICATION NUMBER: FR 89/13770

FILING DATE: 20-OCT-1989

ATTORNEY/AGENT INFORMATION:

NAME: BYRNE, THOMAS E.  
 REGISTRATION NUMBER: 32,205  
 REFERENCE/DOCKET NUMBER: 960-7  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 816-4000  
 TELEFAX: (703) 816-4100  
 TELEX: 200797 NIXN UR  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 557 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-466-974-2

Query Match 3.8%; Score 104.5; DB 2; Length 557;  
 Best Local Similarity 19.3%; Pred. No. 0.26;  
 Matches 114; Conservative 77; Mismatches 150; Indels 251; Gaps 33;

```

QY 18 VVLOGKSSHSSTSRGELAINIRDEPR-----SLDPRGV 52
D 18 VVLOGKSSHSSTSRGELAINIRDEPR-----SLDPRGV 52
D 74 IKLSGQNTSTKCNFSSSLKLVYEIKLIRAKENTSSWYEDSTPRKQIGPPEV 133
QY 53 RLISE-ISLVKHIYGLVQENNLSGNTEPALADYSLSSDGLTYFKLSAFWSNGDPLT 111
D 134 HLEADKKAIVIH-----ISPQKDSVMMALDGLSFTYSL-LIKNS----- 173
QY 112 AEDFISWQVATQVSGIYAFALNPKNVRIQEGHLSIDHGVHSPNSETLVLTLESP 171
D 174 -----SCGE-----ERIENTYSRH-----KIYLSPE 195
QY 172 TSHFLKLA-----LPVPPYKKSQRLQSKSLPIASGAFYRNK----- 212
D 196 TTYCLVKRAALLTSKIGVYSPVHCITVTVENELPP-----PENIEVSQONQYVLKW 248
QY 213 -----OKOMIK--LSKNP--HYNOSOVETKTIITHPIP--ANTAALFNQGL 256
D 249 DYTANMTQVQWMLHAFILKRNPGNHLYKKQ-----IPDCENVKTKQCVFPQ--- 295
QY 257 NMGPFWGERIPQETLSNLOSKG---HLHSPDVAGTSMLEFNINKE-----PLN 302
D 296 -----NVFOKGIYLLRYQASDGNNTSFWESEI-KDTEIQAFLLPYF 337
QY 303 NMLKEALASALDEKALVSTIFLGRAKTADHLIPNHSYP-----EHQOMARQAY 356
D 338 NTR---SLSDSFH-----ITIGAPKQSGN-TPV-IDYDPLIYEIIFWENTSNAERKI 385
QY 357 AKKLFKEALELQITAKDLEHLNLIIPVSSASLSLYQLIREOKESIGRAIPYKKEFA 416
D 386 EKK-----TDVTYVNLKPL-VYCVARAHITM-----DEKLNSSVFSADACEK--- 428
QY 417 LLAADLSSGNFS---LATGGWFADPADPM-----AFLITFAY--PSGVPYAINH--- 461
D 429 -----TRPGNTSKIMLIVGICIALEFALPEVIAAKVFLRCINIVFPPLSKPSSIDEXFS 483
QY 462 -----KDFL-----EILONT-----EEDOHQKRSSELVSOAS 488
D 484 EDPLEKLLSTSEQIEKCFIENISTATVETEINQDEBDEKRYSSQTSQDS 535

```

RESULT 13  
 US-08-471-453-2  
 Sequence 2, Application US/08471453  
 Patent No. 5886153  
 GENERAL INFORMATION:  
 APPLICANT: MOGENSEN, Knud E.  
 APPLICANT: UZE, Gilles  
 APPLICANT: LUMFALLA, Georges  
 APPLICANT: GRESSER, Ion  
 TITLE OF INVENTION: CDNA FRAGMENT CODING FOR THE GENE FOR  
 THE ALPHA INTERFERON RECEPTOR AND PROCESS FOR THE

TITLE OF INVENTION: PREPARATION OF THE CORRESPONDING PROTEIN  
 NUMBER OF SEQUENCES: 2  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: NIXON & VANDERHIVE P.C.  
 STREET: 1100 NORTH GLEBE ROAD  
 CITY: ARLINGTON  
 STATE: VIRGINIA  
 COUNTRY: U.S.A.  
 ZIP: 22201-4714  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/471,453  
 FILING DATE: 06-JUN-1995  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/900,642  
 FILING DATE: 15-JUN-1992  
 APPLICATION NUMBER: FR 89/13770  
 FILING DATE: 20-OCT-1989  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BYRNE, THOMAS E.  
 REGISTRATION NUMBER: 32,205  
 REFERENCE/DOCKET NUMBER: 960-7  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 816-4000  
 TELEFAX: (703) 816-4100  
 TELEX: 200797 NIXN UR  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 557 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-471-453-2

Query Match 3.8%; Score 104.5; DB 2; Length 557;  
 Best Local Similarity 19.3%; Pred. No. 0.26;  
 Matches 114; Conservative 77; Mismatches 150; Indels 251; Gaps 33;

```

QY 18 VVLOGKSSHSSTSRGELAINIRDEPR-----SLDPRGV 52
D 18 VVLOGKSSHSSTSRGELAINIRDEPR-----SLDPRGV 52
D 74 IKLSGQNTSTKCNFSSSLKLVYEIKLIRAKENTSSWYEDSTPRKQIGPPEV 133
QY 53 RLISE-ISLVKHIYGLVQENNLSGNTEPALADYSLSSDGLTYFKLSAFWSNGDPLT 111
D 134 HLEADKKAIVIH-----ISPQKDSVMMALDGLSFTYSL-LIKNS----- 173
QY 112 AEDFISWQVATQVSGIYAFALNPKNVRIQEGHLSIDHGVHSPNSETLVLTLESP 171
D 174 -----SCGE-----ERIENTYSRH-----KIYLSPE 195
QY 172 TSHFLKLA-----LPVPPYKKSQRLQSKSLPIASGAFYRNK----- 212
D 196 TTYCLVKRAALLTSKIGVYSPVHCITVTVENELPP-----PENIEVSQONQYVLKW 248
QY 213 -----OKOMIK--LSKNP--HYNOSOVETKTIITHPIP--ANTAALFNQGL 256
D 249 DYTANMTQVQWMLHAFILKRNPGNHLYKKQ-----IPDCENVKTKQCVFPQ--- 295
QY 257 NMGPFWGERIPQETLSNLOSKG---HLHSPDVAGTSMLEFNINKE-----PLN 302
D 296 -----NVFOKGIYLLRYQASDGNNTSFWESEI-KDTEIQAFLLPYF 337
QY 303 NMLKEALASALDEKALVSTIFLGRAKTADHLIPNHSYP-----EHQOMARQAY 356
D 338 NTR---SLSDSFH-----ITIGAPKQSGN-TPV-IDYDPLIYEIIFWENTSNAERKI 385

```





```

QY 177 KLALPPEFVPHVSQRTQLOKSLP-----IASAFYPK-NIKOKW-KILSKNP-- 223
Dh 1033 KSSAISTSL-TINKLRSSSSQIPPTSTSNQFSASGIFKPSYDELKFPVMDLKVTSVEGC 1092
QY 224 -HYNNOUYERTKIITHIPD-----ANTAA-----KLNOCKLWNOGPPMGKRI 267
Dh 1092 PLVDYAKIQGVLESILYHIEHADTDDJLANNKTFIDITIKIINOEYVO---ENDKRL 1147
QY 268 PQELTSLQSGHLSFDVACTSMWAPENIKF-PLNNKRLREALASALDEKALVSTIFLG 326
Dh 1148 DDLKNNNSQCKRKNCSODNNSAKTFPHVDARPEUSNNKRGATMLGSSIA----- 1201
QY 327 RAKTADHLLPTNHSYPRHQOEADQRAYAKKLRKALBELQITADLEHMLJFPVSS 386
Dh 1202 -----ALEKLOCTLOD-----LVYKIKS 1219
QY 387 SASLLWLOIREQWKESTGFAIPDYKKEFALLQADLSSGNLSLATGSMFADPADPAELIT 446
Dh 1220 S-----YQOLATRLPLGTATRCNKRYVHMLQOPQOTSMALMNG-SSDITLCKM-VT 1268
QY 447 IFAPGVGPPYAINHKDFELICNIOEQEDH--QKRSLEYSQASLILETFHAIIEYDA 504
Dh 1269 EF-----QALKMTDYDIT-NTWYIKLDHFITSKILVSNQD-WIQVSOJLELSNDS 1318
QY 505 ----FOFANMKKLSLGVSPGYVD 525
Dh 1319 LVALFEYPLHAFESNNVIASSSSOLD 1343

```

Search completed: July 26, 2002, 04:38:56  
Job time: 3864 sec



```

2282 AGAAAGTAATGTTGGCACTGTAAAGTGTTCAGATCTTGTCAGATACC 2331
18 ValValLeuGln... GlyCysLysGluSerSerHisSerThrSerat 33
:::|||||:::|||||:::|||||:::|||||:::|||||:::
2332 GCTTGGCAGCTTGTGTATATAAAGTTCAGAGGCTCAGAAAGTACC 2381
33 gglgluLeuAlaIleasnIleargaspIuProargSerLeuaspProa 50
2382 AAAAAGCTAACAGTACGTTTGTGTAAAGTTCAGAAATCATGTGATTATA 2431
50 rglInValArgLeuLeuSerGluIleSerLeuValLysHisIleTyrGlu 66
2432 TTTTGACTAATGGCGGTGAGACGATGTGATACACAAATGGTTGAT 2481
67 GlyLeuValGlnGluAsnAsnLeuSerGlyAsnIleGluProAlaLeuAla 83
|||||:::|||||:::|||||:::|||||:::|||||:::
2482 GGTCTTTG... GAAAGCATGAGTATGTTATTTAGTACCATGACTTGC 2528
83 agLysPtyrSerLeuSerSeraspIleuThrTyrThrPheLysLeuL 100
|||||:::|||||:::|||||:::|||||:::|||||:::
2529 TAAAGATGGAGGTTTCAAGACGCTGACTTATCTATCTACTTTC 2578
100 ySerAlaPhe... TrpSerAsnGlyAsp... ProLeu 110
|||||:::|||||:::|||||:::|||||:::|||||:::
2579 GCGATGCTGTCTCTTGGTATACCGCTGATGAGAAATATGCCCGATG 2628
111 ThrAlaGluaspPheIleGluSerTyrPlyGlnValAlaThrGlnGlu 127
|||||:::|||||:::|||||:::|||||:::|||||:::
2629 ACAGCAGAAAGATTTGTGACTGGTTGAGACACGCGGTTCAGTAATAC 2678
127 lSerGlyIleTyrAlaPheAlaLeuAsnProIleLysAsnValArgLysI 144
|||||:::|||||:::|||||:::|||||:::|||||:::
2679 AGATGCTCTTACCTGTGTA... GATTCATTAATAAAGCTTAAGGCTT 2725
144 leGlnGlnGlyHisLeuSerIleaspHisPheGlyValHisSerProas 160
|||||:::|||||:::|||||:::|||||:::|||||:::
2726 ACCAAAAGGTGAGTAGATTTTAAAGAGTGGTGTAAAGCCCTTGAC 2775
161 GluSerThrLeuValValThrLeuGluSerProThrSerHisPheLeu 177
|||||:::|||||:::|||||:::|||||:::|||||:::
2776 GATAAAGCTGTTCAGTACTTGAACAGCCCTGAAGCTACGTAATTC 2825
177 sleuLeuAlaLeuProValPhePheProValHisLysSerGlnArgThr 194
:::|||||:::|||||:::|||||:::|||||:::|||||:::
2826 AAAAAGCACTTATAGTGTCTTCCCACTT... AATGCCAAATTTT 2869
194 euGlnSerLys... SerLeuProIle 201
|||||:::|||||:::|||||:::|||||:::|||||:::
2870 TGAAGCTAAAAGTAAAGATTTTGTACACCGATCCATCATCAATCCT 2919
202 AlaSerGlyAlaPheTyrProLysAsnIleLysGlnLysGlnTrpIle 218
|||||:::|||||:::|||||:::|||||:::|||||:::
2920 GTTATGTGCTTACTTCTTGAGCCGCTTCACTCAATCAATCATATGGA 2969
218 sleuSerLysAsnProHisTyrTyrAsnGlnSerGlnValGluThrLys 235
:::|||||:::|||||:::|||||:::|||||:::|||||:::
2970 ATTCATTAATAAAGTAAAGTACGATGATCAAGATGTTGGATGAT 3019
235 hTrIleThrIleHisPheIle... ProaspAlaAsnThrAlaAlaLys 249
:::|||||:::|||||:::|||||:::|||||:::|||||:::
3020 CTGTAAATGACTTACTCAGATGTGTCAAGCCAGCTGTCTCTACAG 3069
250 leuPheAsnGlnGlyLysLeuAsn... TrpIleGlnLysPro 262
|||||:::|||||:::|||||:::|||||:::|||||:::
3070 AACTTTGACAGAGGTAGTTCAGCGCTTCACGACTTATACCAATATACC 3119
262 oTrrpGlyArgIleProGlnGluThrLeuSerAsnLeuGlnSerLysG 279
:::|||||:::|||||:::|||||:::|||||:::|||||:::
3120 TACCTACAAATCAGCTAAGAAATATATCCTGATTAAT... 3159
279 LysHisLeuHisSerPheaspValAlaGlyThrSerThrPheLeuThrPheAsn 295
|||||:::|||||:::|||||:::|||||:::|||||:::
3160 ...ACTTACGAATGTGACTGAGATATCCCTCATTTACATGGAAT 3204

```

```

296 lIleAsnLysPheProLeuAsnMetLysLeuArgGluAlaLeuAlaSe 312
:::|||||:::|||||:::|||||:::|||||:::|||||:::
3205 TTGAACCGTAAGTCTTTCATAAAGCACTAAGAAAGACCTGCACAAACA 3254
312 rAlaLeuAspLysGlnAlaLeuValSerThrIlePheLeuGlyArgAlaL 329
|||||:::|||||:::|||||:::|||||:::|||||:::
3255 TGCC... GGTAAAGAAAGCTTTTACACAAAGATTTT... 3289
329 yStrAlaAspHisLeuLeuProThrAsnIleHisSerTyrProGlnHis 345
:::|||||:::|||||:::|||||:::|||||:::|||||:::
3290 GTCAAGCTATTCAGTGTCTTTGACCGAGCGCTATTC... AAGAC 3333
346 GlnLysGln... GluMetAlaGlnArgGlnAlaTyrAlaLysLeuP 361
:::|||||:::|||||:::|||||:::|||||:::|||||:::
3334 AAATGAGAGTCAGATGTCACAAACAAAGCTTACGTATACATGCTGTG 3383
361 elyGlnAlaLeuGlnGluLeuGlnIleThrAlaLysAspLeuGlnHisL 378
:::|||||:::|||||:::|||||:::|||||:::|||||:::
3384 CCACCAAGCTTTGTACCGTTGAG... 3408
378 euAsnLeuIlePheProValSerSerSerAlaSerSerLeuValGln 394
:::|||||:::|||||:::|||||:::|||||:::|||||:::
3409 ... 3426
395 LeuIleArgGlnGlnTrpLys 401
3427 AAGTTGAAAAGCAATGGCNA 3447
seq_name: /cgn2_6/ptodata/2/ina/5B.COMB.seq:us-08-245-511-46
seq_documentation_block:
? Sequence 46, Application US/08245511
? Patent No. 5928900
? GENERAL INFORMATION:
? APPLICANT: Masure, H Robert
? APPLICANT: Pearce, Barbara J
? APPLICANT: Tuomanen, Elaine
? TITLE OF INVENTION: BACTERIAL EXPORIED PROTEINS AND
? TITLE OF INVENTION: CELLULAR VACCINES BASED THERON
? NUMBER OF SEQUENCES: 58
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Klauber & Jackson
? STREET: 411 Hackensack Avenue
? CITY: Hackensack
? STATE: New Jersey
? COUNTRY: USA
? ZIP: 07601
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/245,511
? FILING DATE: 18-MAY-1994
? CLASSIFICATION: 424
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/116,541
? FILING DATE: 01-SEP-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: Jackson Esq., David A.
? REGISTRATION NUMBER: 26,742
? REFERENCE/DOCKET NUMBER: 600-1-069 CIP
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 201 487-5800
? TELEFAX: 201 343-1684
? TELEX: 133521
? INFORMATION FOR SEQ ID NO: 46:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2019 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: both

```



```

1429 GACTGG.....GATTTATCATAT.....1446
430 aThGlyGlyTyrPheAlaSPheAlaSPrometaAlaPheLeu....445
1447 .AATGCGTGGGGTCACAGCTTGGCGATCATCACTACCTGATA1455
446 .....ThiIlePheAlaIlyPro451
1496 TCATCAACCATCTGTAGAGAAAGTACTAAACATATTAGGTTTAC1545
452 SerGly.....ValProTyrAlaIleAsnHisLysAsPheLe465
1546 TCAGGGGAGATTAATGTAGCTGTAAAAAGTAGTCTATATGACTACGA1595
465 uGluIleLeuGlnAsnIle...GluGlnGlnGlnAspHisGlnLysArgS481
1596 AAAATGTGTTACTGAGGCTGTGTATGAGACTACAGATGTTGCTAAACGCT1645
481 erGluLeuValSerGlnAlaSerLeuTyrLeuGluThrPheHisIleLe497
1646 ATGATAAATAGCTGCACGCCACACTTGTTCACAGATATGCTTTGATT1695
498 GluPro499
1696 ATTCCA1701

seq_name: /cgn2_6/prodata/2/lna/5B_COMB.seq:US-08-600-993A-46
seq_documentation_block:
; Sequence 46 Application US/08600993A
; Patent No. 5981229
; GENERAL INFORMATION:
; APPLICANT: Masure, H Robert
; APPLICANT: Pearce, Barbara J
; APPLICANT: Tuomanen, Elaine
; TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESS: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/600,993A
; FILING DATE: 1-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/245,511
; FILING DATE: 18-MAY-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/116,541
; FILING DATE: 01-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-069 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2019 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; STRAIN: R6
; IMMEDIATE SOURCE:
; CLONE: SPUR98
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1932
; US-08-600-993A-46

alignment_scores:
Quality: 270.50 Length: 602
Ratio: 0.902 Gaps: 25
Percent Similarity: 49.834 Percent Identity: 21.595

alignment_block:
US-09-824-567-2 x US-08-600-993A-46 ..
Align seg 1/1 to: US-08-600-993A-46 from: 1 to: 2019

19 ValLeuGlnGlyCysLysGluSerSerHisSerThSerArgLysI 35
||||| :|||: |||: |||:
4 GTACTGCACAGACTCTGTGATCAGGTTCAACGCTTAAGGTGAGAAAC 53
||||| :|||: |||: |||:
35 uLeuAlaIleAsnIleArgAspGluProArgSerLeuAspProArgLysI 52
||||| :|||: |||: |||:
54 ATTCTATACATTATTAGACAGACCTGTATACCTCAACTATTGACAA 103
||||| :|||: |||: |||:
52 alArgLeuLeuSerGluIleSerLeuValLysHisIleTyrGluLys 68
||||| :|||: |||: |||:
104 CTCTAAGCTCGACAGCAAAATTTACAGTAAAGTGTGTTGATGTTG 153
||||| :|||: |||: |||:
69 ValGlnGluAsnAsnLeuSerGlyAsnIleGluProAlaLeuAlaGlu 85
||||| :|||: |||: |||:
154 CTA..GAAATGATCGCTACGGAGAACTTGTCCGCTATGCGTGAGGA 200
||||| :|||: |||: |||:
85 pTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLysLeu...Lys 101
||||| :|||: |||: |||:
201 TTGCTGTATATCCAGATGATGATTGACTTACCTTATCTATCCGTAG 250
||||| :|||: |||: |||:
101 erAlaPheTrp.....SerAsnGlyAsp.....ProLeuThAla 112
||||| :|||: |||: |||:
251 ATGCATAATGATATCTTGAAGGTGAGATACCGCGCATGAAGCT 300
||||| :|||: |||: |||:
113 GluAspPheIleGluSerTrpLysGlnValAlaThrGlnGluValSer 129
||||| :|||: |||: |||:
301 CAAGACTTTGTACAGAGACTAAATATGCTGCTATAAAATACAGATCG 350
||||| :|||: |||: |||:
129 yIleTyrAlaPheAlaLeuAsnProLysAsnValArgLysIleGln 146
||||| :|||: |||: |||:
351 TCTTACCT...GTCAACAATCAATCAAGGTTGATGCTTATGTAA 397
||||| :|||: |||: |||:
146 lGlyHisLeu...SerIleAspHisPheGlyValHisSerProAsnGlu 161
||||| :|||: |||: |||:
398 AAGGGGAATCAAGATTTCTCAACAAGTAAAGGCTGCGTGA 447
||||| :|||: |||: |||:
162 SerThrLeuValIleThrLeuGlnSerProHisSerHisPheLeuLys 178
||||| :|||: |||: |||:
448 CAGAGGTTCAGTACACTTGAACAACAGAAAGCTTGTGAATTC 497
||||| :|||: |||: |||:
178 uLeuAlaLeuProValPhePheProValHisLysSerGlnArgThrLeu 195
||||| :|||: |||: |||:
498 GACAACCATGGGTGCTGTCGCCAGTT...AATGAGAGATTTTGA 541
||||| :|||: |||: |||:
195 lserLys.....SerLeuProIle 201
||||| :|||: |||: |||:

```

```

542 ATTCAAAAGAGATGATTTTCCAAAGCTACGATCAAGATGATCTCTTG 591
202 AlaSerGlyAlaPheTyrProIysAsnIleGlyGlnTyrPheIle 218
592 TATAAGGCTCTTATTTGTAATCAATCTGACCAATCTCTGTTGA 641
218 sLeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValIleThrLys 235
642 ATTGGCAAAATCCGAACACTGCTGGATAGACCAATGTCATATTGCA 691
235 hIleThrIleHisPheIle.....ProAspAlaAsnThrAlaAlaLys 249
692 AAGTTAAATTCTCATCTCGGATGCTCAGATACAGCAAAACCTGGAGAA 741
250 LeuPheAsnGlnGlyLysLeuAsnTyrGlnGlyProProTyrGlyLys 266
742 AACTTTAAAGATGCTACCTTACA.....GCAGCTCG 773
266 gIle..... 267
774 TCTCTATCCAAAGATGCAAGTTTGGCAGACTGACAAAGATATGAGG 823
268 .....ProGlnGluThrLeu.....Ser 273
824 ACAATATTGCTATCTACTCAACAAGACTTATACGTATCTAGTGGTACA 873
274 AsnLeuGlnSerLysGlyHisLeuHisSerPheAspValIleGlyHis 290
874 AATATTGACCGCTACGCTCTATAATACACA.....TCTAAGACCG 914
290 TTTTLeuThrPheAsnIleAsnLysPheProLeuAsnMetLysLeuA 307
915 CGATCAACAAGAGCATCACTCAAAAGCTCTCTTAAACAAGATTTCC 964
307 TGGIleAlaLeuAlaSerAlaLeuAspLysGlnAlaLeuValSerThrIle 323
965 GTCAAGCTATTGCTTGGTTGATCTTACAGCCCTATGCTCTCAGTTG 1014
324 .....PheLeuGlyTyrAlaLysThrAlaAspHisLeu.....Pr 336
1015 AATGCAACAACCTGACAGCATTAATCTTGCTATCTCTTTGTGCACC 1064
336 cThrAsnIleHisSer..... 341
1065 AACATTTGTCACAGATGTAAAAAATTGGCGATATGTCAAAAGAGA 1114
341 ..... 341
1115 AATTGTCACCTATGCGGATGATGCAAGATGTTAATCTTCAGATTCT 1164
342 .....TyrProGlnHisGlnLysGlnGlnMetAlaGlnAr 353
1165 CAGGATGCTCTTACAAATCCAGAAAAGCCAGCTGATTTGCTAAAGC 1214
353 gGlnAlaTyrAlaLysLysLeuPheLysGlnAlaLeuGlnGln 370
1215 T.....AAATCAGCTTACAGACAGCAAGAGTG 1240
370 LeThrAlaLysAspLeuGlnHisLeuAsnLeuIlePheProValSer 386
1241 TGACATTCCCA.....ATTCAATTTGGATATGCCAGTGCACAG 1278
387 SerAlaSerSerLeuLeu.....ValGlnLeuIleArgGlnGlnTyr 401
1279 ACAGCAACTACAAAGATTCAGCGGCTCAATCATGAACAATCTCTGGA 1328
401 sGlnSerLeuGly..... 405
1329 AGCAACTTTAGAGCTGATATGTCATTAATGATATTCACAACACTACAA 1378
406 .....PheAlaIleProIleValGlyLys 413
1379 AAGAGCAAGTAACAATATTACATATTTCIGATAAATGCTGCTGGCGAA 1428

```

```

414 GluPheAlaLeuLeuGlnAlaAspLeuSerSerGlyAsnPheSerLeuA 430
1429 GACGCG.....GATTTATCAGAT..... 1446
430 aThrGlyGlyTyrPheAlaAspPheAlaAspPheAlaPheLeu.... 445
1447 AATGCGGTGGGTCCAGACTTGGCCGATCCAGCAACCTACCTGTGTA 1495
446 .....ThrIlePheAlaTyrPro 451
1496 TCATCAACCATCTGTAGAGAAAGTACTAAACAATATTAGGTTGAC 1545
452 SerGly.....ValProProTyrAlaIleAsnHisLysAspPheLe 465
1546 TCAGGGGAAGATAATGTGCTGCTAAAAAAGTAAAGCTATATAGCTAGA 1595
465 uGlnIleLeuGlnAsnIle...GlnGlnGlnGlnAspHisGlnLysArg 481
1596 AAATTTGTTACTGAGGCTGGTGTATGAGACTACAGATGTTGCTAAAGCT 1645
481 eGlnLeuValSerGlnAlaSerLeuTyrLeuGluThrPheHisIleIle 497
1646 ATGATTAATTCGCTGACGCCCAAGCTTGGTTGACAGATATGCTTGTATT 1695
498 GluPro 499
1696 ATTCCA 1701

```

```

seq_name: /cgn2_6/prodata/2/ina/6B.COMB.seq:US-09-103-840A-2
seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: TUBERCULOSIS
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

```

```

alignment_scores:
Quality: 176.00 Length: 528
Ratio: 0.649 Gaps: 21
Percent Similarity: 51.326 Percent Identity: 20.265

```

alignment\_block:

US-09-824-567-2 x US-09-103-840A-2/rev ..

Align seg 1/1 to reverse of: US-09-103-840A-2 from: 1 to: 4403765

```

38 ILeAsnIleArgAspGluProArgSerLeuAspProArgGlnValArgLe 54
4099170GTCAACGGCGGGGAACGCCCAACCCGCTGATCCAGACGGGACCAAGA 4099121
54 uLeuSerGluIleSerLeuValLysHisIleTyrGlnGlyLeuValGlnG 71

```





alignment\_scores:  
 Quality: 176.00 Length: 528  
 Ratio: 0.649 Gaps: 21  
 Percent Similarity: 51.326 Percent Identity: 20.265

alignment\_block:  
 US-09-824-567-2 x US-09-103-840A-1/rev ..

Align seg 1/1 to reverse of: US-09-103-840A-1 from: 1 to: 4411529

```

38 IleasnlleargaspGluProargSerLeuAspProargGlnValArgLeu 54
   ::::: ::::: ::::: ::::: ::::: :::::
4106737GTCAACGGCGGGGACCGCCCAACCGGCTGATCCGACCGCGCACACAA 4106924

54 uLeuSerGluLeuSerLeuValLysHisIleTyrGluGlyLeuValGln 71
   ::::: ::::: ::::: ::::: ::::: :::::
4106923CACGACGAGTGGCGGCGATGATGATGCTTTCGCGCGCTGATGCTCT 4106874

71 LuAsnAsnLeuSerGlyAsnIleGluProAlaLeuAlaGluAspTyrSer 87
   ::::: ::::: ::::: ::::: ::::: :::::
4106873ATGACGCGCGTGGC.....AGCGGTGCTGGAGGTGCGCGCAGTCA 4106833

88 Leu...SerSerAspGlyLeuThrTyrThrPheLysLeuLysSerAlaPh 103
   ::::: ::::: ::::: ::::: ::::: :::::
4106832ATTCGAGAGCCCGATACGTCAATACGATGCTGCTCAACCCGCGCTG 4106783

103 e...TyrSerAsnGlyAspProLeuThrAlaGluAspPheIleGluSer 119
   ::::: ::::: ::::: ::::: ::::: :::::
4106782AAATTACCGAGCGCTGCCCGGTGACGCGCCACTGCTTGTGTCAGCG 4106733

119 rPlySglnValAlaThrGlnGluValSerGlyIleTyrAlaPheAlaLeu 135
   ::::: ::::: ::::: ::::: ::::: :::::
4106732GGGACTACGAGCGCTGAGACCAATGCGCACTGCAGACGACTTTTTC 4106683

136 AsnProIleLysAsnValArgLysIle.....GlnGluGln 147
   ::::: ::::: ::::: ::::: ::::: :::::
4106682AGCCCGATCGAAGGTTGATGACGTGCGCGCGCGCGGAGTAATAAG 4106633

147 yHisLeuSerIleAspHisPheGlyValHisSerProAsnGluSerTrp 164
   ::::: ::::: ::::: ::::: ::::: :::::
4106632CCGACCATGTC.....GGTGTCCCGCTGTCACACGACTGGAGT 4106589

164 euValAlaThrLeuGluSerProThrSerHisPheLeuLysLeuAla 180
   ::::: ::::: ::::: ::::: ::::: :::::
4106588TTCACGCTTGAAGCGCGGACATCGACTGACGTGGGCTTGGC 4106539

181 LeuProValPhePheProValHisLysSerGln...ArgTrpLeuGln 196
   ::::: ::::: ::::: ::::: ::::: :::::
4106538CAGAGCTGTTTATCCGCTGCGGACTCGCATTTTGGGACATGGCGCG 4106489

196 rLysSerLeu...ProIleAlaSerGlyAlaPhe..... 206
   ::::: ::::: ::::: ::::: ::::: :::::
4106488GTTCGCGCGCATCGGACATCGGCGCGCGCGGCAACAATCCGCCAC 4106439

207 .....TyrProLysAsnIleLysGlnLysGlnTrpIleLys 218
   ::::: ::::: ::::: ::::: ::::: :::::
4106438CAGCGCGCGCGCTGGGAACAACAAGTCA.....ATGAC 4106401

219 LeuSerLysAsnProHisTyrTyrAsnGlnSerGlnValGluTrpLys 235
   ::::: ::::: ::::: ::::: ::::: :::::
4106400CTGTGCGCCAAACCCGACTACACGCGCAACCGCAACCGCGGTAACA 4106351

235 rIleThrIleHisPheIleProAspAlaSerThrAlaAlaLysLeuPhe 252
   ::::: ::::: ::::: ::::: ::::: :::::
4106350TTTGGAGTTCAGTTCAGCGCAATTCGACACCGCTATGCGGACTGCT 4106301

252 singlyLysLeuAsnTrpGlnGly.....ProTrpTrpGlyGluArg 266
   ::::: ::::: ::::: ::::: ::::: :::::
4106300TGTCTCGCAATCTGATGTCGAGACAGATTCGCGCGAGCGGTGAGC 4106251

267 IleProGlnGluThrLeuSerAsnLeuGlnSerLysGlyHisLeuHis 283
   ::::: ::::: ::::: ::::: ::::: :::::

```

```

4106250GTCTACACGCGGACCTGGCGACCAAGTACAGCGGCGCGACGAT 4106201
283 rPheAspValAlaGlyThrSerTrpLeuThrPheAsnIleAsnLysPhe 300
   ::::: ::::: ::::: ::::: ::::: :::::
4106200CAATTCAGACCTTCGACACT.....CCGTTGGCGGTGCGCATTTGC 4106160

300 rOleuAsnAsnMetLysLeuArgLys...AlaLeuAlaSerAlaLeuAsp 315
   ::::: ::::: ::::: ::::: ::::: :::::
4106159GTGCGAGAGGCGCGGTGCGCGGTGCGGTGCGGTGCGGTGCGGTG 4106110

316 LysGlnAlaLeuValSerThrIlePheLeuGlyArgAlaValThrAla 331
   ::::: ::::: ::::: ::::: ::::: :::::
4106109CGCGCGCAAAATCTCGACAGATCTTCGCGGAGACCCGACATTCG 4106060

332 .....AspHisLeuLeuProThrA 338
   ::::: ::::: ::::: ::::: ::::: :::::
4106093GATTTACACCCCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 4106010

338 sn...IleHisSerTyr...ProGlnIleGlnLysGlnMetAlaGln 352
   ::::: ::::: ::::: ::::: ::::: :::::
4106009ACGAGTGTCTGACTACAGCCGCGACGCGGCTCGGCGGCTGAGG 4105960

353 ArgGlnAlaTyrAlaLysLysLeuPheLysGlnAlaLeuGlnGluLeu 369
   ::::: ::::: ::::: ::::: ::::: :::::
4105959CGCGATCGCATCTACCGGTGAGCGCGCGGTACGCGGATC..... 4105921

369 nIleThrAlaLysAspLeuGlnHisLeuAsnLeuIlePheProValSer 386
   ::::: ::::: ::::: ::::: ::::: :::::
4105920.GCCATACCGCTACCGCTGCGGTGCGGTGCGGTGCGGTGCGGTG 4105872

386 eSerAlaSerSerLeuLeu...ValGlnLeuIleArgGlnTrpLys 401
   ::::: ::::: ::::: ::::: ::::: :::::
4105871ACAGCATCAAGACGTTGGGATCGATGGGCTGCGCGCGCGCGAC 4105822

402 GluSerLeuGlyPheAlaIleProIleValAlaLysLysLysLeuLeu 418
   ::::: ::::: ::::: ::::: ::::: :::::
4105821ACCTTTCGCGGATTTTCGACCCGATCAACACCGCGCATCGAC.... 4105777

418 uGlnAlaAspLeuSerSerGlyAsnPheSerLeuAlaThrGlyLysTrp 435
   ::::: ::::: ::::: ::::: ::::: :::::
4105776.....AGTGCATTTTCGCGGTGATGAGC 4105755

435 heAlaAspPheAlaAspProMetAlaPheLeuThr...IlePheAlaTyr 450
   ::::: ::::: ::::: ::::: ::::: :::::
4105754GGGCGACTACCGCTGATGATGATGATGATGATGATGATGATGAT 4105705

451 ProSerGlyValProProTyrAlaIleAsnHisLysAspPheLeuLys 467
   ::::: ::::: ::::: ::::: ::::: :::::
4105704GGCGCGGATCCAGACGCTGCGCTACATCAACCGGAAATTCGACG 4105655

467 eLeuGlnAsnIleGluGlnGlnGlnAspHisGlnLysArgSerGluLeu 484
   ::::: ::::: ::::: ::::: ::::: :::::
4105654CTGCGCGCGCGCGGCGGCGGCGCGCGCGGCGGCGGCGGCGG 4105605

484 aLysGlnAlaSerLeuTyrLeuGluThrPheHis.....IleIleGlu 498
   ::::: ::::: ::::: ::::: ::::: :::::
4105604TCAACGACGCGCGCAAGCAATCTG.....TTTCACACATATGCA 4105561

499 ProIleTyrHisAspAlaPheGlnPheAlaMetAsnLysLysLeuSer 515
   ::::: ::::: ::::: ::::: ::::: :::::
4105560CGCGCTCGGATTTACATGATGCTGCTGATGCTGCTGCTGCTGCT 4105511

515 nLeuGlyLysSerProThrGlyValAlaLysPhe 526
   ::::: ::::: ::::: ::::: ::::: :::::
4105510GTACACGCTACCTCGATGTCTGCGCGGTGCG 4105477

seq_name: /cgn2_6/prodata/2/lna/6B.COMB.seq:US-09-385-028-20
seq_documentation_block:
; Sequence 20, Application US/09385028
; Patent No. 6232106
; GENERAL INFORMATION:
; APPLICANT: Susan E. Jensen

```

APPLICANT: Kwamena A Aidoo  
 APPLICANT: Ashish S. Parthakar  
 TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic  
 Patent No. 6232106

TITLE OF INVENTION: Acid Biosynthesis

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC

STREET: The Jenner Building, 400 Seventh Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/385,028

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/790,462

FILING DATE: 29-JAN-1997

ATTORNEY/AGENT INFORMATION:

NAME: D. Douglas Price

REGISTRATION NUMBER: 24,514

REFERENCE/DOCKET NUMBER: 1418/P57452US2

TELEPHONE: (202) 638-6666

TELEFAX: (202) 39305350

TELEX: RCA 248593 IDEA UR

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 1668 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-09-385-028-20

#### alignment\_scores:

Quality: 165.50 Length: 462

Ratio: 0.710 Gaps: 17

Percent Similarity: 50.433 Percent Identity: 20.996

#### alignment\_block:

US-09-824-567-2 x US-09-385-028-20 ..

Align seg 1/1 to: US-09-385-028-20 from: 1 to: 1668

34 GlyGluLeuAlaIleAsnIleArgAspGluProArgSerLeuAspProArg  
 85 GGGAGCGTGGCGGTGCGCGACGGAGCGACTCGACTCGCGCGCGCGG 134  
 50 gGlnValArgLeuLeuSerGluIleSerLeuValLysHisIleTyrGluG 67  
 135 CAACACGTACTACGCGCTACACCTGGAACCTTCCTCGGCTCATGCGCGGA 184  
 67 LLeuValGlnGluAsnAsnLeuSerGly.....AsnIleGlu 79  
 185 CGCTGGTACCTTCGACACCGCGCGGAGAGCGCGCGCGCGCTCGG 234  
 80 ProAlaLeuIleGluAspTyrSerLeuSerSer...AspIleuThrTy 95  
 235 CCGGACCTCGCGCGAGTGGCGCGAGTCCGAGAGAGCGCGCGCTG 284  
 95 ThrPheLysLeuLysSerAlaPhe...TipSerAsnGlyAspProLeuT 111  
 285 GACCTACCGGCGTGGCGAGGCGCTGCGCTACAGAGAGCGCGCGCTG 334

111 hrlaGluAspPheIleGluSerTrpLysGlnValAlaThrGlnGluVal 127  
 335 TCTGGCGCGCATC.....AAGCGCGCATGCGCGCGAGCA 372  
 128 SerGlyIleTyrAlaPheAlaLeuAsnProIleLysAsnValArgLysI 144  
 373 TACGGACCGAGTGTCTGGCGCGCGTCC.....AC 404  
 144 eGlnGluGlyHisLeuSerIleAspHisPheGly..... 155  
 405 CTACTTCGCGCATCTCGGCGCGAGTACGGCGCGCGCGCGGAGC 454  
 156 .....ValHisSerProAsnGluSerThrLeu 164  
 455 CGAGCGCGCGAGCGCGGTGACCTGAGAGCGCGCGCGAGCGCGCTG 504  
 165 ValValThrLeuGluSerProThrSerHisPheLeuLysLeuAlaLe 181  
 505 GTCTTCGCGCTGCGGAGCGCGTTCGCGGAGTGTCTGCGCGCGCAT 554  
 181 uProValPhePheProValHisLysSerGlnArgThr...LeuGlnSer 197  
 555 GCGGTCCACACCGCGCGCGCGCGCGAGCGCGCGCGCGCGAGTACC 604  
 197 ySerLeuProIleAlaSerGlyAlaPheTyrProLysAsnIleLysGln 213  
 605 GCGTCCGCGCGCGTGGCGCGCGCGCGTACCGGATCTGTCTGACCCG 654  
 214 LysGlnTrpIleLysLeuSerLysAsnProHisTyrTrpAsnGlnSer 230  
 655 GCGCGCTGGCGCGTCTGGAGCGCGCATCGCGCGCGAGCGCGCGCGA 704  
 230 n.....ValGlnThrLysThrIleThrIleHisPheIlePro 243  
 705 CCGGTCGCGCTCCAGCGCGCGCTCCGCGATGAGTGCACCTGCGCAAG 754  
 243 spAlaAsnThrAlaLysLeuPheAsnGlnGlyLysLeuAsnTrpGln 259  
 755 ACCCGACAGAGTGGACCGCATGCTGCTGGCGCGCGCGCGCGCATG 804  
 260 GlyProProTrpGly.....GlnArgIleProGln 270  
 805 CTCGCGCGTTCGCTGGCGCGCGCGCGCGCGCGCGCGCGATC 846  
 270 uThrLeuSerAsnLeuGlnSerLysGlnHisLeuHisSerPheAspVal 287  
 847 .....CTGCGCGAGCGCGAGCTGCGCGCGCGCGCGCGCGCG 889  
 287 laGlyThrSerTrpLeuThrPheAsnIleAsnLysPhe...ProLeuAsn 302  
 890 CCGGCTTCACCTGAGTACTGCTGCGAGCGCGGATCGCGCGGTTCGAC 939  
 303 AsnMetLysLeuArgGlnAlaLeuAlaSerAlaLeuAspLysGlnAlaLe 319  
 940 AAGTGCACCTGCGCGCGCGCGTGCAGTTCGCGCGCGCGCGCGAT 969  
 319 uValSerThrIlePheLeuGlnArgAlaLysThrAlaAsp.....H 333  
 990 GCGAGAGCGCG.....TAGCGCGCGCGGTGGCGCGCGCGCGCGACA 1033  
 333 LsLeuLeuProThrAsnIleHisSerTyrProGlnHisGlnLysGln 349  
 1034 CCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1083  
 350 MetAlaGlnArgGlnAlaTyrAlaLysLysLeuPheLysGlnAlaLeu 366  
 1084 GTGGCG.....CCGAGGCGCGCGCG 1103  
 366 uGluLeuGlnIleThrAlaLysPheGlnHisLeuAsnIlePhePhe 383  
 1104 CAGCTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1153  
 383 roValSerSerSerAlaSerSerLeuLeuValGlnLeuIleArgLysGln 399

```

1154 GCTTCGCGACGATCGCCGCCGACAGGACCGCTACAGGATACCGG 1203
400 TrpLysGlnSerLeuGlyPheAlaIleProIleValGlyLysGlnPheAl 416
1204 GCCGCGGAGCGCTGGCGCGGCTCGCCGCGGCTCGGCGATCGAGCGCGA 1253
416 AleuLeuGlnAlaAspLeuSerGlyAsnHe..... 427
1254 GGTGCTG.....CACTTCCCTGTGGGCGACTACTTCGACCGCTACGCGG 1297
428 .....SerLeuAlaThrGlyGly 433
1298 GCTGCCGCGAGTATCTCGCGACGACGAGGATCGGATCATCATCTTCGCG 1347
434 TrpPheAlaAspPheAlaAspProMetAlaPheLeu 445
1348 TGGGGCGCGCACTTCCCGCGACGAGATACGCGCTTCCTC 1383

```

seq\_name: /cgn2\_6/prodata/2/lna/6B\_COMB.seq:US-09-385-028-13

seq\_documentation\_block:

```

; Sequence 13, Application US/09385028
; Patent No. 6232106
; GENERAL INFORMATION:
; APPLICANT: Susan E. Jensen
; APPLICANT: Kwamena A. Aidoo
; APPLICANT: Ashish S. Parakekar
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6232106
; TITLE OF INVENTION: Acid Biosynthesis
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, P/C
; STREET: The Jennifer Buliding, 400 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,028
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/790,462
; FILING DATE: 29-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: D. Douglas Price
; REGISTRATION NUMBER: 24,514
; REFERENCE/DOCKET NUMBER: 1418/P57452US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-6666
; TELEFAX: (202) 39305350
; TELEX: RCA 248593 IDEA UR
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11604 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-385-028-13

```

alignment\_scores:

```

Quality: 165.50 Length: 462
Ratio: 0.710 Gaps: 17
Percent Similarity: 50.433 Percent Identity: 20.996

```

alignment\_block:

US-09-824-567-2 x US-09-385-028-13 ..

Align seq 1/1 to: US-09-385-028-13 from: 1 to: 11604

```

34 GlyLeuAlaIleAsnIleArgAspGluProArgSerLeuAspProAr 50
US-09-824-567-2 ..
Align seq 1/1 to: US-09-385-028-13 from: 1 to: 11604
7293 GGGACCGCTCGGCTGTCCGACGAGGACGACTTCGATCGCTGACCCCGG 7342
50 GlnValArgLeuLeuSerGluIleSerLeuValIleTyrGlnG 67
7343 CAACAGCTACTAGCTTACACTGGAATCTTCGCGCTATCGGCCGGA 7392
67 LyeuValGlnGluAsnAsnLeuSerGly.....AsnIleLeu 79
7293 CGGTGTACCTTCGACACCGCGCGGCAAGCGCGGCGGCTGTG 7442
80 ProAlaLeuAlaGlnAspTyrSerLeuSerSer...AspGlyLeuThrTy 95
7443 CCCGACTTCGCGAGTCTGTGGGAGTCTCTCCGAGAGCGCGCGGTCTG 7492
95 rThrPheLysLeuLysSerAlaPhe...TrpSerAsnGlyAspProLeu 111
7493 GACCTACGCGGCTGCGCGAGGCTCGGCTACGACGACGACGCGCGGTG 7542
111 hrAlaGluAspPheIleGluSerTrpLysGlnValAlaThrGlnGluVal 127
7543 TCTCGCGCGACATC.....AACACGCCATTCGCCCGACGACAC 7580
128 SerGlyIleTyrAlaPheAlaLeuAsnProIleLysAsnValArgLysI 144
7581 TAGGCGACCGATGCTGTGGCGCGGCTGCG.....AC 7612
144 eGlnGlnGlyHisLeuSerIleAspHisPheGly.....AC 7612
7613 CTACTTCGCGACCTCTGTGGACGACGACGCGCGCGCGGAGAC 7662
156 .....ValHisSerProAsnGluSerThrLeu 164
7663 CGGACGCGGACGACGCGGTACCGCTGAGACCCCGACGACGACGCGTG 7712
165 ValValThrLeuGlnSerProThrSerHisPheLeuLysLeuAlaLe 181
7713 GCTTCCCGCTGCGGAGCGCTTCCGCGGAGATGCTGTGGCGACAT 7762
181 uproValPhePheproValHisLysSerGlnArgThr...LeuGlnSer 197
7763 GCCGTCCACACCGCGCTGCGCGGACCGGACACCGCGCGGAGTACC 7812
197 ysserLeuProIleAlaSerGlyAlaPheTyrProLysAsnIleLysGln 213
7813 GCGTGGGCGCGTGGCGACCGCGCTACCGGATCGTCTGTACACCGG 7862
214 LysGlnTrpIleLysLeuSerLysAsnProHisTyrTyrAsnGlnSer 230
7863 GCGAGCTGGCGCTCTGAGCCCAATCCGACGAGACCGGACCGGACGA 7912
230 n.....ValGlnThrLysThrIleThrIleHisPheIleProA 243
7913 CCCGTCGCGGCTCAAGCGCGCTCCGATGAGTGCACCTCGGCAAG 7962
243 spAlaAsnThrAlaAlaLysLeuPheAsnGlnGlyLysLeuSerTrpGln 259
7963 ACCGCGACGAGTGGACCGCATGCTGCGGCGGACGAGCCCATGTGGAC 8012
260 GlyProProTirpely.....GluArgIleProGlnI 270
8013 CTCGCGGCTTCGTGTGACCGCGCGGCGGAGAGGCGATC..... 8054
270 uThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerPheAspValA 287
8055 .....CTGGCGAGCCGAGACTGCGCGGACGAGGACAC...CCGCTGA 8097

```

```

? CLASSIFICATION: ?
? PRIOR APPLICATION DATA: ?
? APPLICATION NUMBER: US 08/790,462
? FILING DATE: 29-JAN-1997
? ATTORNEY/AGENT INFORMATION:
?   NAME: D. Douglas Price
?   REGISTRATION NUMBER: 24,514
?   REFERENCE/DOCKET NUMBER: 1418/P574S2US2
? TELECOMMUNICATION INFORMATION:
?   TELEPHONE: (202) 638-6666
?   TELEFAX: (202) 39305350
?   TELERX: RCA 248593 IDEA UR
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
?   LENGTH: 15079 base pairs
?   TYPE: nucleic acid
?   STRANDEDNESS: single
?   TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEetical: NO
? ORIGINAL SOURCE:
? ORGANISM: Streptomyces clavuligerus
?
? US-09-385-028-1
?
?
? Alignment_scores:
?     quality: 165.50      Length: 462
?     Ratio: 0.710        Gaps: 17
? Percent Similarity: 50.433    Percent Identity: 20.996
?
? Alignment_block:
? US-09-824-567-2 x US-09-385-028-1 ..
?
? Align seg 1/1 to: US-09-385-028-1 from: 1 to: 15079
?
? 34 GlyGlnLeuAlaIleAsnIleArgSpaGluProAlaSerLeuAspProAr 50
? ||| ||| :::: ||||| ||||| |||||
? 9325 GGGACCGCTCGCGGTGTCGCCAGCAGCACTTGCAGTGGCTGCACCCCGG 9378
?
? 50 gclValAlrAglLeuLserGluIleSerLeuValLyshIsIlEtyGlUG 67
? ::::::
? 9375 CAACAAGACTACCCTCACACTGGAACCTTCCTCGGCTCATCGGGCGGA 9428
?
? 67 lYdeuValGIInGLIAsnAsnLeuSerGly.....AsnIleGlu 79
? ||||| :::::: ||||
? 9425 CGGTGTACCTTGACACACCGCGCGGAGAAGCGCGGCTGTG 9474
?
? 80 ProAlaIeuAlaGluAspTrYSerLeuSer...AspGlyLeuThrTy 95
? ||| ||||| :::: ||||| ||||| ::::::
? 9475 CCGGACCTCGCGGAGTGTGCGGCGAGTCTCCAGAGACCGCGGTTCTG 9524
?
? 95 rThrPheLysIeuLysSerAlaPhe..TTPSerAsnGlyAspProLeuT 111
? :||:::||:::||:::||:::||:::||:::||:::||:::||:::||
? 9525 GAACCTACGGGTGGGAGAGGGCTCGCTACAGAGAGCGAGCGCGTGTG 9574
?
? 111 hPaIaGluAspPheIIeGluSerTrpLYsglNalAlaThrGlnIleVal 127
? :::: ||||| :::::: |||||
? 9575 TCTCGGCGGACATC.....AAgCAGCGCATCGCGCAGCAAC 9612
?
? 128 SerGlyIleLyrlAlaPheAlaLeuAsnProIleLysAsnValArgLYsII 144
? ||| :::::: ||::||
? 9613 TAGGACACCATGCTCTCGGCGCGCGGTGCG.....AC 9644
?
? 144 eGlNGluGlyHISLeuSerLIeaSPHisPhetly..... 155
? |||||
? 9645 CTACTTCGCGACCTCTCTGGGACCGGAGTACCGGCGCCCTCGGCGGAGC 9694
?
? 156 .....ValISerProksnGluSerThreU 164
? 9695 CGAGCGCGAGACGACCGGTGAAGCGTGGAGACCCCGAGACGAGCAGCTGTG 9744
?
? 165 ValValThrLeuGlnSerProThrSerHisPheLeuLysIeuAlaLea 181
? ||| ||||| ||||| ||||| |||||

```

```

9745 GTCCTCCGGCTGCGGGAGCCGTTCCGGGGATGATGCTGCGGACCAT 9794
181 uProValPhePheProValHisLysSerGlnArgThr...LeuGlnSerL 197
:::|||||:::|||||:::
9795 GCGGTCACACACCCCGCTGCCCGGACCGGACACCGGCGCGGATACC 9844
197 ySSerLeuProIleAlaSerGlyAlaPheTyrProLysAsnIleLysGln 213
:::|||||:::|||||:::
9845 GAGTCGGCGCCCGTGGCGACCGCCGTCACCGGATGCTGTACACCCCG 9894
214 LysGlnTrrPileLysLeuSerLysAsnProHisTyrAsnGlnSerG 230
:::|||||:::|||||:::
9895 GCGGAGCGCGCTGCTGTGAGCCCAATCCGACCTGGAGCCCGACCGCA 9944
230 n.....ValGluThrLysThrIleThrIleHisPheIleProA 243
:::|||||:::|||||:::
9945 CCGCGTCGCGTCACACCGCGCTCCCGGATGAGCTGACACCTCGCAGG 9994
243 sPAlaAsnThrAlaAlaLysLeuPheAsnGlnGlyLysLeuAsnTrrP 259
:::|||||:::|||||:::
9995 ACCCGCACAGGTGAGACCCCATGCTGCTGGCGGCGCAGCCCATGTGAG 10044
260 GlyProTrrPgly.....GluArgIleProGlnG 270
:::|||||:::|||||:::
10045 CTCGCGGCGCTGCGTGTGAGCCCGCGGCGCCAGAGCGCATC..... 10086
270 uThrLeuSerAsnLeuGlnSerLysGlyHisLeuHisSerPheAspValA 287
:::|||||:::|||||:::
10087 ....CTGCGGAGCCGAGACTCCGCGCACCGGCGACAC...CCGTGA 10129
287 laGlyThrSerTrrPleuThrPheAsnLeuAsnLysPhe...ProLeuAsn 302
:::|||||:::|||||:::
10130 CCGGCTTCACTGATGATCTACTGCTGCGAGCGGATCCCGCTTGCAC 10179
303 AsnMetLysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGlnAlaLe 319
:::|||||:::|||||:::
10180 AATGTGACTGCCGCGGCGGCTGCAGTCCGACCGCACCAAGCGCGCAT 10229
319 uValSerThrIlePheLeuGlyArgAlaLysTrrAlaAsp.....H 333
:::|||||:::|||||:::
10230 GCMAGAGGCG.....TACGGCGGCGGCGGCGGCGGACATCGGACCA 10273
333 IsLeuLeuProThrAsnIleHisSerTyrProGlnHisGlnLysGlnG 349
:::|||||:::|||||:::
10274 CCGCTGCGGCGCGGACCTTCACAGGCTACACAGACTTCGACCGTACCCG 10323
350 MetAlaGlnArgGlnAlaTyrAlaLysLysLeuPheLysGluAlaLeuG 366
:::|||||:::|||||:::
10324 GTCGCG.....CCGAGAGGACCGG 10343
366 uGluLeuGlnIleThrAlaLysAspLeuGlnHisLeuAsnIlePheP 383
:::|||||:::|||||:::
10344 CGACTGTGAGCGCGCGCGCGGAGCTGAGGCTGCGCGGATGCCGACG 10393
383 roValSerSerSerAlaSerSerLeuLeuValGlnLeuIleArgGluGln 399
:::|||||:::|||||:::
10394 GCTTCGACACAGATGCGCGCCGCGACAGACCGGCTCAAGAGATACCGG 10443
400 TrrPylsGlnSerLeuGlyPheAlaIleProIleValGlyLysGluPheAl 416
:::|||||:::|||||:::
10444 GCGGCGGAGCGCTGCGCGGCGCTGCGCGGCTGCGCATCGAGGCGGA 10493
416 aleuLeuGlnAlaAspLeuSerSerGlyAsnPe..... 427
:::|||||:::|||||:::
10494 GTGTCTG.....GACTTCGCGTGGCGGACTACTTCGACCGCTACGCGG 10537
428 .....SerLeuAlaThrGlyGly 433
:::|||||:::|||||:::
10538 GCTGCCCGGAGATCTGCGGACGACGAGGATGCGGATATATGTTGGC 10587
434 TrrPheAlaAspPheAlaAspProMetAlaPheLeu 445
:::|||||:::|||||:::
10588 TGGGGCGCGGACTTCCCGACGAGATACGCGCTTCTC 10623

```

```

seq_name: /cgn2_6/prodata/2/1na/PCrUS_COMB.seq:PCT-US96-05320A-113
seq documentation block:
; Sequence 113, Application PCrTUS9605320A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences
; APPLICANT: 9410 Key West Avenue
; APPLICANT: Rockville, MD 20850
; APPLICANT: United States of America
; APPLICANT: Johns Hopkins University
; APPLICANT: 720 Rutland Avenue
; APPLICANT: Baltimore, MD 21205
; APPLICANT: United States of America
; APPLICANT: Mark D. Adams
; APPLICANT: Owen White
; APPLICANT: Hamilton O. Smith
; APPLICANT: J. Craig Venter
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20003-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/05320A
; FILING DATE: April 22, 1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: June 7, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: June 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Eric K. Steffe
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.014PC01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2540
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2232 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; PCT-US96-05320A-113

alignment_scores:
Quality: 138.50 Length: 654
Ratio: 0.478 Gaps: 32
Percent Similarity: 44.343 Percent Identity: 18.807

alignment block:
US-09-824-567-2 x PCT-US96-05320A-113 ..
Align seg 1/1 to: PCT-US96-05320A-113 from: 1 to: 2232
6 ValGlyIleCysIleThrIleLeuLeuSerLeuSerValIleGlnG 22
:::|||||:::|||||:::
1 ATGATATTTTAAATTATTAACGATATCTTTTATTGATTACCTTGAGG 50
22 yCysLys.....GluSerSerHisSerSerThrSerArgG 34

```



906	CACATATGTACTGAAAAATTAACCATGGATTAACTATTCTACACG...	951
255	ysLeuAsnTrpGlnGlyProProTrpGlyGluatGllleProGlnGluTrh	271
952	...TTGAGCTACAGTAAACACACAAATGATGCTGCCAAGAGCA	999
272	LeuSerAsnLeuGlnSerLysGlyHisLeuHisSerPheAspValaIagI	288
1000	TCTTACGATTACGGGCT...ACCTTAGAGTAA	1028
288	yTrhSerTrpLeuThrPheAsnIleAsnLysPheProLeuAsnAsnMetL	305
1029	TCAAGGCTGCACTACTATTACAGACTCTACTTTGATATAAATAC...	1074
305	ysLeuArgGluAlaLeuAlaSerAlaLeuAspLysGluAlaLeuValaSer	321
1075	...ACAAAGCTTTTAATATCAAAACTGCTTCAT	1107
322	ThrIlePheLeuGly...	326
1108	GAACTATATATTGGTTGGCATGGTTAAAAACAGAAATACCTTAAAT	1157
327	...ArgAlaLysThrAlaAspHisLeuLeuProThrAsnI	339
1158	GTATGATGAAGTAAAGTAAATGAGAAACCGAT...	1188
339	LeuHisSerTrpProGlnHisGlnLysGlnLumetalaglnatglnala	355
1189	...TATAATTAAGCGTATTATTCACAACTTATATGCAATCGTGCACGTCA	1236
356	TrpAlaLysLysLeuPheLysGluAlaLeuGlnGluLeuGlnIleThrAl	372
1237	TATACACGCGCGTTTATTATTACA...	1260
372	AlaLysPheGlnLumHisLeuAsnLeuIlePheProValSerSerSerAla	389
1261	CATATAAATTAATGACAGCATATATATTTTAGCACACGCA...	1299
389	erSerLeuLeuValGlnLeuIleArgGlnGlnTrpLysGlnSerLeuGly	405
1300	...GTGCGTTATGACCATCATCATATATATATAGCG	1329
406	PheAlaIleProIleValGlyLysGluPheAlaLeuLeuGlnAlaAspIe	422
1330	...CAGAAAAATTTTACCCTGATAAATATATATATAT	1361
422	uSerSerGly...	433
1362	CTCTGACGACAAATGATATTATAGCCAAAAATTTAAT...	1399
433	LYTTPPheAlaAspPheAlaAspProMetAlaPheLeuThrIlePheAla	449
1400	GTGG...	1407
450	TrpProSerGlyValProProTrpAlaIleAsnHisLysAspPheLeuG	466
1408	TATTTATTTAAGTTTA...AATATGATGTAAACCAT...	1448
466	uIleLeuGlnAsn...	474
1449	TTTATTATTCAGATTTTATGTAATAACTTGGCGACACCTGTATATGATGANC	1498
474	LugInAspHisGlnLysAlaGSerGlyLeuValaSerGlnAlaSerLeuTr	490
1499	AGTATGAGACGCAATTTAAGCAATCTTGTGCGCTGACACCTCTTTAAAT	1548
491	LeuGlu...	493
1549	TTAGAAAAAAGATGATTAATATAACAGAGAGTGGGCGGATATATATCTCT	1598
493	rPheHisIleIleGluProIleTrhHisAspAlaPheGlnPhe...	507
1599	CAATCATCTATTTTCAGGA...	1642

```

508 .....AlaMetAsnLysLysLeuSerAsnLeuGlyVal 518
1643 CTATTTTAAATCCCGGCAAGATGAATATTAAACCAAGAGAGTT 1692
519 SerProthrgly 522
1693 AATTGTGTAGGC 1704

seq_name: /c9n2_6/plodata/2/lna/6B_COMB.seq:US-08-462-467B-3
seq documentation block:
; Sequence 3, Application US/08462467B
; Patent No. 6210899
; GENERAL INFORMATION:
; APPLICANT: Rosenbaum, Jan S
; TITLE OF INVENTION: The use of a BMP protein Receptor
; TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells
; TITLE OF INVENTION: Co-Transfected With a Type II BMP Receptor and a Type I
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; STREET: The Procter & Gamble Company
; CITY: Ross
; STATE: OH
; COUNTRY: USA
; ZIP: 45061
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,467B
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Herstro, Bart S
; REGISTRATION NUMBER: 32,572
; REFERENCE/DOCKET NUMBER: 5474R
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 627-0633
; TELEFAX: (513) 627-0260
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2156 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-462-467B-3

alignment_scores:
Quality: 117.00 Length: 485
Ratio: 0.544 Gaps: 24
Percent Similarity: 44.330 Percent Identity: 21.031

alignment block:
US-09-824-567-2 x US-08-462-467B-3 rev ..

Align seq 1/1 to reverse of: US-08-462-467B-3 from: 1 to: 2156

4 115SerValGlyIleCysIleThrIle.....LeuLeuSe 15
|||||
1982 ATGACGTTGACATTTGATGATGGGCTCACAGATTGTTCTTTC 1933
15 r1euserValValLeuGlnGlyCysLysSerHisSerHis 32
|||||
1932 CCAATCATCATTAAGTTCAGCAATCCCTTCTCAGACACAGTGCAGTAA 1983
32 e1rargLylGluLeuAlaIleAsnIleArgAspLupProArgSerLeuasp 48

|||||
1882 GCCGAGCCTTCATCCTGCTGCCAACAGCTTCGATTGCTCCTAGC 1833
49 ProArgGlnValArgLeuLeuSerGlnIleSerLeuValLysHisIleTy 65
1832 GACCTCACTGCCAGGCTATTTCCT..... 1809
65 rGlnGlyLeuValGlnGluAsnLeuSerGlyAsnIleGluProAlaI 82
1808 .....TTCAGGCTTCTGCAACTTG..... 1788
82 eValGluAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLys 98
1787 .....GCTCTGTTTTCCTAGAC 1767
99 LeuLysSerAlaPhe.....TrpSerAsnGlyAspProLeuThr 112
1766 ACGAAGACCTGATATCTCCAAAAGGAGATGTTCCAACTCTGCTG 1717
112 lAgLAspPheIleGluSerTrpLysGlnValAlaThrGlnGluValSer 128
1716 AAAAG.....CCATCTGACTCTGTGACGATTCCTCCGGAAGA 1676
129 GlyIleTyrAlaPheAlaLeuAsnProIleLysAsnValArgLysIleG 145
1675 GCTCTGATCATCATTAATATCTCC.....AATGATTATAGTCA 1635
145 nGluGlyHisLeuSerIleAspHisPheGlyValHisSerProAsnGluS 162
1634 AAGAGATCATCTACTGTTTCAAAAGCTGATTCACAGCTCCCTCAAGTT 1585
162 eThrLeuValValThrLeuGluSerPro..... 171
1584 CACAGCTCTCTGACAGCACTTCTGTCCTATATCTATATGCTCCACCT 1535
172 .....ThrSerHisPheLeuLysLeuAlaLeuProValPheHepr 186
1534 CGCTTATGCGCATATATCTCTCCCTCGGCGCACAGTATATTTC.. 1487
186 oValHisLysSerGlnArgThrLeuGlnSerLysSerLeuProIleAlaIS 203
1486 .....CAGTCAGCTCATGACAGTCAAA 1462
203 eT.....GlyAlaPheTyrPro 208
1461 GTCCTAATAACACAGTTCATCATTTTTCATAGACATTTCTGCTGT 1412
209 LysAsnIleLysGlnLysGlnTrpIleLysLeuSerLysAsnProHisTy 225
1411 TTAAATCTCATGGGAATTGCAGGTT..... 1385
225 TYrAsn..... 227
1384 TATAATGATCTCTGCTGTAATCTGTGAGAGATAACCCAGTCTCT 1336
228 .....GlnSerGlnValGluThrLysThrIleThrIleHisPheIlePro 242
1335 AGTAACAGAAATGAGCAAGGCAAGACCTTACCCAGCTGCTGTGGA 1286
243 AspAlaAsnThrAlaIleLysLeuPheAsnGlnGlyLysLeuAsnTrpG 259
1285 GACTTAATACTTGCAATA.....AAGATCCATTGGGA 1254
259 nGlyProTrpTrpGlyGluArgIlePro.....GlnGluThrLeuS 273
1253 TAGACTTCATTCACAGCAAAATATATCCATGGTCCATCTGCACTACT 1204
273 e1rAsnLeuGln..SerLysGlyHisLeuHisSerPheAspValAlaGlyT 289
1203 CTCATCTCAACTATTAACGGGCAATGTTGTCATGTCATCAATAAGCA 1154
289 h1rSerTrpLeuThrPheAsnIle..... 296

```

```

1153 CTCGTAAATGTTCTTTTCGTTGATAAATTCGTGAGGTTTGCAAGGAA 1104
297 ASnlySpheProLeuAsnAsnMetLysLeuArgLysLeuAlaSerAl 313
||| ||| :||| :||| ||| :|||
1103 AACACTTTTACAGAACTGACGCTCAGCTCAGCAAGAGGCTTTATATCTCC 1054
313 aleuAspLysGluAlaLeuValSer.....ThrIlePheLeuGlyArgA 328
| :||| :||| :||| :|||
1053 TCCATATCGACCTCGCCAAATCAGCTCCACAGATTTCAGATTATCTAGAT 1004
328 lAlYsThrAlaAspHisLeuLeu..... 335
:|||||
1003 CAAGAGAGGTTGGATGCTGCTGCCTCATCATGTTCACTACTGTGAGAA 954
336 .....ProThrAsnIleHisSerTyrProGlnHisGln.. 346
|||||:|||||
953 CCGTTGTTACGGTCTCCTGCTCAACT...CTGATCCAAAGATATAGGC 907
347 .....LysGlnGluMetAlaGlnArgGlnAlaTyrAlaLysLysLeuP 361
||| :||| :||| :|||
906 AACATATCAAAACAGCTAATACAGACAGTGCACCAAGCAATGATTATG 857
361 helYsGluAlaLeuGluGlu.....LeuGln.IleThrAlaLysAs 374
:|||||:|||||
856 TCTATCTCGCTTAAATGAGTAGAGTGCAGCTGAGTGTGTGTGCAGCA 807
374 pLeuGlu.....HisLeuAsnIleLeuPheProV 384
:|||||:|||||
806 GGTGAGAAATTCACAGTAAAGTTGACATTACATTAATCTGTGCTAC... 761
384 aLSeSerSerAlaSerSerLeuValGlnLeuIleArgGlnInfrp 400
:|||||:|||||
760 .....AACAGCAAGAAAGGTATGTTCCATTCGTGATGAGGAGAGAGTGG 716

seq_name: /cgn2_6/ptdata/2/ina/6b_COMB.seq:US-08-334-179A-3
seq_documentation_block:
? Sequence 3 Application US/08334179A
? Patent No. 6306622
? GENERAL INFORMATION:
? APPLICANT: ROSENBAUM, JAN S.
? APPLICANT: NOHNO, TSUTOMU
? TITLE OF INVENTION: CDNA ENCODING A BMP TYPE II RECEPTOR
? NUMBER OF SEQUENCES: 14
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: THE PROCTOR AND GAMBLE COMPANY
? STREET: 11810 EAST MIAMI RIVER ROAD
? CITY: ROSS
? STATE: OH
? COUNTRY: US
? ZIP: 45061
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: PatentIn Release #1.30, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/334,179A
? FILING DATE: 04-NOV-1994
? CLASSIFICATION: 435
? ATTORNEY/AGENT INFORMATION:
? NAME: CORSTANTIE, BRAHM J
? REGISTRATION NUMBER: 34,804
? REFERENCE/DOCKET NUMBER: 5473
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 513-627-2858
? TELEFAX: 513-627-0260
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2157 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear

```

```

? MOLECULE TYPE: CDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: join(409..2154)
US-08-334-179A-3

alignment_scores:
  Quality: 117.00      Length: 485
  Ratio: 0.544
  Percent Similarity: 44.330      Percent Identity: 21.031

alignment_block:
US-09-824-567-2 x US-08-334-179A-3/rev ..

Align seg 1/1 to reverse of: US-08-334-179A-3 from: 1 to: 2157

4 lIleSerValGlyIleCysIleThrIle.....LeuLeuSe 15
||||| ||| :|||||
1982 ATAGCAGTAGACATTGGATGACTGTGGGCTCAGAGTTGTTCTTTC 1933
15 rIleuSerValValLeuGlnGlyCysLysGluSerSerHisSerThrS 32
| :||| :||| :||| :|||
1932 CCAAATCATCATTAAGTTCAGCCATCCTTCCACAGACACATGTCAGTAA 1883
32 eRArgGlyGluLeuAlaIleAsnIleArgAspGluProArgSerLeuAsp 48
|||||:|||||
1882 GCCGAGCCTTCGATCTCGTCCCAAGAGCTCTTCGATTGTCCTTGAGT 1833
49 ProArgGlnValArgLeuLeuSerGluIleSerLeuValLysHisIleTyr 65
:|||||:|||||
1832 GACCTCATCTGCCAGGCTATTTC..... 1809
65 rGluGlyLeuValGlnGlnAsnAsnLeuSerGlyAsnIleGluProAla 82
:|||||:|||||
1808 .....TTCAGGCTCTCGGAGACTTG..... 1788
82 euAlaGluAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLys 98
|||||:|||||
1787 .....GGTCTGTGTTTCCCTAGAC 1767
99 LeuLysSerAlaPhe.....TrpSerAsnGlyAspProLeuThrA 112
:|||||:|||||
1766 ACGAAGACCTGCATATCTCCAAAGAGGAGTGTTCACACCTCTGTGCG 1717
112 lAgLysPheIleGlnSerTyrPylsGlnValAlaThrGlnGluValSer 128
:|||||:|||||
1716 AAAAG.....CCATCTGTACTGTGTAAGGAGATTCCCTGGAGAGA 1676
129 GlyIleTyrAlaPheAlaLeuAsnProIleLysAsnValArgLysIleG 145
||||| ||| :|||||
1675 GGTCTGTACATCTCATTAATATCTCC.....AATGATTAAGTCA 1635
145 nGluGlyHisLeuSerIleAspHisPheGlyValHisSerProAsnGluS 162
||| :||| :||| :|||
1634 AGAGCAATACATGTCCTACTGTTTCAAGCTGATTCACAGCCCTCAAGT 1585
162 eThrIleValValThrLeuGluSerPro..... 171
| :||| :||| :||| :|||
1584 CACAGCTCTTTCATAGCACTGTGGCCATATATCATGATAGTCCAACT 1535
172 .....ThrSerHisPheLeuLysLeuValAlaLeuProValPhePhe 186
||||| ||| :|||||
1534 CGCTTAGGCTGCATTAATCTCTCCCTCGGCGCAACAGTCAATTC.. 1487
186 oValHisLysSerGlnArgThrLeuGlnSerLysSerLeuProIleAlaS 203
||||| ||| :||| :|||
1486 .....CACTCAGCCTCATGTGACAGTCCAAA 1462
203 eR.....GlyAlaPheTyrPro 208
|||
1461 GTCACTATATACACAGGTTCCATCATTTTTCACATGACATTTTGTGCTGT 1412

```

```

209 LysAsnIleLysGlnLysGlnIleThrIleLysLeuSerLysAsnProHisTy 225
    |||...|||...
1411 TTAATCTCGATGGAAATTGCAGGTT..... 1385
225 rTyrAsn..... 227
    |||...
1384 TATAATGATCTCCCTCGTGAATTCTGTGGAAGATTAAGCCAGCTCCT 1336
228 .....GlnSerGlnValGluThrLysThrIleThrIleHisPheIlePro 242
    |||...|||...|||...
1335 AGTAACGAGATGAGCAAGACGCAAGACTTACCACGACTCATCTGTGTGGA 1286
243 AspAlaSerThrAlaAlaLysLeuPheAsnGlnGlyLysLeuAsnTrpG1 259
    |||...|||...|||...
1285 GACTTAATTAATCTGCATA.....AAGATCAATGGGA 1254
259 ngIleProProTrpGlyGluArgIlePro.....GlnGluThrLeus 273
    |||...|||...|||...
1253 TAGTACTCCATCAACAAGCAATATTCATGCGTCACCTGCAGTGCCT 1204
273 eArAsnLeuGln..SerLysGlyHisLeuHisSerPheAspValAlaGlyT 289
    |||...|||...|||...
1203 CTCACTCCCACTATAAAGCGGCAATGTTGTCATGTCATCAACAAGCA 1154
289 hrSerTrpLeuThrPheAsnIle..... 296
    |||...|||...|||...
1153 CTCTGTAATGCTCTTTCCTGTTGATAAATTCGACGGTTGCCAAAGCA 1104
297 AsnLysPheProLeuAsnAsnMetLysLeuArgLysAlaLeuAlaSerAl 313
    |||...|||...|||...
1103 AACACTTTTCAGACACTGAGACGCTCATCCAAAGACCTTTATATACGCG 1054
313 aLeuAspLysGluAlaLeuValSer.....ThrIlePheLeuGlyArg 328
    |||...|||...|||...
1053 TCATATTCGACCTCGGCCAATCACTGCATCAACAGTTTCAGATTATCTGAT 1004
328 lAlaThrAlaAspHisLeuLeu..... 335
    |||...|||...|||...
1003 CAAGAGAGGGTTCGAGTCTGCTCCATCATCATGTCATCACTGAGAGA 954
336 .....ProThrAsnIleHisSerTyrProGluHisGln.. 346
    |||...|||...|||...
953 CCTGTTTACGGTCTCTCCGTCAACATT...CTGTATCCAAAGCAATAGGC 907
347 .....LysGlnGluMetAlaGlnArgAlaTyrAlaLysLysLeuP 361
    |||...|||...|||...
906 AACTATCAAAACAGCTAATACAGAGCTGATGCCAAAGCAATGATTATTC 857
361 heLysGluAlaLeuGlnGlu.....LeuGln..IleThrAlaLysAs 374
    |||...|||...|||...
856 TCCATCTCCGTTAAATGAATGAGTGCATGAGTGTGTTGTGTGCGA 807
374 pleuGlu.....HisLeuAsnLeuIlePhePro 384
    |||...|||...|||...
806 GGTGAAATTCCTAGTAAGTTGACATTACATAAATCTGTGCTAC.... 761
384 aIleSerSerAlaSerSerLeuLeuValGlnLeuIleArgGlnGlnTrp 400
    |||...|||...|||...
760 .....AACAGCAAAAGCGTATCTTCATCTCGAATTAGGAGAGGTGG 716

```

seq\_name: /cgr2.6/ptodata/2/ina/6B.COMB.seq:US-08-334-179A-1

```

seq_documentation_block:
/ Sequence 1, Application US/08334179A
/ Patent No. 6306622
/ GENERAL INFORMATION:
/ APPLICANT: ROSENBAUM, JAN S.
/ APPLICANT: NOHNO, TSUTOMU
/ TITLE OF INVENTION: CDNA ENCODING A BMP TYPE II RECEPTOR
/ NUMBER OF SEQUENCES: 14
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: THE PROCTER AND GAMBLE COMPANY
/ STREET: 11810 EAST MIAMI RIVER ROAD

```

```

/ CITY: ROSS
/ STATE: OH
/ COUNTRY: US
/ ZIP: 45061
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent Release #1.30, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/334,179A
/ FILING DATE: 04-NOV-1994
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: CORSTANJE, BRAHM J.
/ REGISTRATION NUMBER: 34,804
/ REFERENCE/DOCKET NUMBER: 5473
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 513-627-2858
/ TELEFAX: 513-627-0260
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 3601 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: join(409..3522)
/ US-08-334-179A-1

```

```

alignment_scores:
Quality: 117.00 Length: 485
Ratio: 0.544 Gaps: 24
Percent Similarity: 44.330 Percent Identity: 21.031
alignment_block:
US-09-824-567-2 x US-08-334-179A-1/rev ..

```

Align seq 1/1 to reverse of: US-08-334-179A-1 from: 1 to: 3601

```

4 lIleSerValcIlyTlEcysIleThrIle.....LeuLeuSe 15
    |||...|||...|||...
1982 ATACAGTAGACATATGATGATGATGCGCTGCACAGATTGTCTTCTTC 1933
15 lLeuSerValcIlyLeuGlnGlyCysLysGluSerSerHisSerThrS 32
    |||...|||...|||...
1932 CCAATATCATATATAGTTTACAGCACTCTTCTCAGCACACTGTCGATGA 1883
32 eArArgLysLeuValAlaIleAsnIleArgAspGluProArgSerLeuAsp 48
    |||...|||...|||...
1882 GCGAGCCTCTGATCCTCGGTGCGCCACAGCTTCGTGATGTTCTCTCTGAGT 1833
49 ProArgGlnValArgLeuLeuSerGlnIleSerLeuValLysHisIleTy 65
    |||...|||...|||...
1832 GACCTCAGTCCAGGCTATTTCT..... 1809
65 rGluGlyLeuValGlnGlnAsnLeuSerGlyAsnIleGluProAlaL 82
    |||...|||...|||...
1808 .....TTCACGGCTTGTGGCAACTGG..... 1788
82 euAlaGluAspTyrSerLeuSerSerAspGlyLeuThrTyrThrPheLys 98
    |||...|||...|||...
1787 .....GGTCTCGTTTTCCTCTGAC 1767
99 LeuLysSerAlaPhe.....TrpSerAsnLysAspProLeuThr 112
    |||...|||...|||...
1766 ACAGAACCTCGATATCTCCAAAGTGGAGTGTTCACACCTCTGCTGT 1717
112 lAgLysPheIleGluSerTrpLysGlnValAlaThrGlnGluValSer 128
    |||...|||...|||...

```

```

1716 AAAAG.....CCATCTGCTACTCTGTCGATTCGCGATTCCTCCGCGAACA 1676
1729 GYIIEYRALPheALeUasNProlIeLysaINValArgLysIleGI 145
1675 GGTCTGACATTCATTAATATCTCC.....AATGAGTATGTCOA 1635
145 nGIUGIYHISLeuSerIleASPHISPhEgIYAlHISserPROASNGIUS 162
1634 AGAGCATACATGCTACTGTTCTCAAGCTGATTCACAGTCCCTCAAGTT 1585
162 erThrLeuValThrLeuGInserPro..... 171
1584 CACAGCTCCCTTAGACACTTCGGGCCATATATCTGATAGGCCAACCT 1535
172 .....ThSerHisPheLeuLysLeuAlaLeuProValPhePhePr 186
1534 CGCTTATGCGTGCATTAATCTTCCTCCCTCGCGCACAGCTATTTCC.. 1487
186 oValHISLysSerGInArgThrLeuGInserLysSerLeuProIleAlas 203
1486 .....CAGTCAGCCTCATGACAGCCCAA 1462
203 er.....GLValPheThyPro 208
1461 GTCACATAATAACACAGTTCATATTTTCATAGACATTTTCCTGCT 1412
209 LysASnIleLysGInLysGInTPPLeLysLeuSerLysAsnProHISLys 225
1411 TTTAAATCTCGATGGAAATTCACAGTT..... 185
225 TTYAsn..... 227
1384 .TATAATGATCTCCCTGCTGTAATTCGTCGTAAGATTAAGCAGCTCTCT 1336
228 .....GInSerGInValGInThrLysThrIleThrIleHisPheIlePro 242
1335 AGTAACAGATAGACAGACGCAAGACGCTTACCCAGTCTTGTGGGA 1286
243 ASPAlaasnThrAlaAlaLysLeuPheasngInGlyLysLeuAsnTrpGI 259
1285 GACTTAATACTTCGATA.....AAGATCCATTGGGA 1254
259 nGlyProPTrpGlyLysIleArgIlePro.....GInGInThrLeuS 273
1253 TAGTACTTCATCACACAAATATTCATGCGCTCATCTGACAGTCTCT 1204
273 eRAsnLeuGIn..SerLysGInHISLeuHisSerPheAspValAlaGIYT 289
1203 CTCATCTCCAACTATTAAGCGGCAATGTTTCATGTTCCATCAAGGCA 1154
289 hISerTrpLeuThrPheAsnIle..... 296
1153 CTCTGTAATATGTTCTTCTGCTGATTAATTCATGACGTTTGCAAGGAA 1104
297 AsnLysPheProLeuAsnAsnMetLysLeuArgIleAlaLeuAlaSerAl 313
1103 AACACTTTTACAGCACTGACGCTCATCAAGAGCCTTATATATCTCTC 1054
313 AlEuaSPlySGInAlaLeuValSer.....ThrIlePheLeuGlyArgA 328
1053 TTCATATTCGACCTCGCGCAATTCATCAACAGTTTCAGATTATCTAGAT 1004
328 lAlYsThrAlaAspHISLeuLeu..... 335
1003 CAAGAGAGGTTGCGATGCTGCTGCTCATCATGATTCATCTACTCGAAGA 954
336 .....ProThrAsnIleHisSerTyProGInHISGIn.. 346
953 CTTTGTTCAGCTCTCTCTCAACAT.....CTGATATCAACACATAAGC 907
347 .....LysGInGInMetAlaGInArgGInAlaIYzAlaLysLysLeuP 361
906 AACTATCAAAACAGCTAATACAGACTGATGCAAGCAATGATTATTG 857

```

seq\_name: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:US-08-462-467B-1

seq\_documentation\_block:

; Sequence 1, Application US/08462467B

; Patent No. 6210899

; GENERAL INFORMATION:

; APPLICANT: Rosenbaum, Jan S

; TITLE OF INVENTION: The use of a BMP Protein Receptor

; TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells

; TITLE OF INVENTION: Co-Transfected With a Type II BMP Receptor and a Type I

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: The Procter & Gamble Company

; STREET: 11810 East Miami River Road

; CITY: Ross

; STATE: OH

; COUNTRY: USA

; ZIP: 45061

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/462,467B

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Hersko, Bart S.

; REGISTRATION NUMBER: 32,572

; REFERENCE/DOCKET NUMBER: 5474R

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (513) 627-0633

; FAX: (513) 627-0260

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3603 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-462-467B-1

alignment\_scores:

Quality	117.00	Length	485
Ratio	0.544	Gaps	24
Percent Similarity	44.330	Percent Identity	21.031

alignment\_block:

US-09-824-567-2 x US-08-462-467B-1/rev ..

Align seg 1/1 to reverse of: US-08-462-467B-1 from: 1 to: 3603

4 lIeSerValGlyIleCysIleThrIle.....LeuLeuS 15

1982 ATAGCAGTAGACATGATGATGCTGAGCTCACAGATTTGTTCTTC 1933

15 lIeSerValValLeuGInGlyLysGInserThISerSerThIS 32





```

NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586.90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Klebsiella pneumoniae
US-08-743-637B-8

Alignment_scores:
Quality: 113.50 Length: 87
Ratio: 2.027 Gaps: 3
Percent Similarity: 64.368 Percent Identity: 34.483

Alignment_block:
US-09-824-567-2 x US-08-743-637B-8/rev ..

Align seq 1/1 to reverse of: US-08-743-637B-8 from: 1 to: 238

92 GlyLeuThrTyrThrPhelysLeuLysSerAlaPheTrpSerAsnGlyAs 108
|||||
238 GGCTGACC.....TGGTCTGACGGGCAC 216

108 PProLeuThrAlaGluAspPheIleGluSerTrpLysGlnValAlaThrG 125
|||||
215 GGCCATTACAGCGCAGATGCTGTGAGCTGGCAACGGCTGCTCGGC 166

125 InGluValSerGlyIleTyrAla...PheAlaLeuAsn...ProIleLys 139
|||||
165 CGGCACACACCTCCCTACGCCAGCTACCCGGCAATATGCAATATCGCC 116

140 AsnValAlaGlySIIeGlnGluGlyHisLeuSerIleAspHisPheGlyVa 156
|||||
115 AACGCCCGGAGATTGCCCTCGGGCAAAAGGAGACCCGAGAGCTGGGGGT 66

156 IHisSerProAsnGluSerThrLeuValAlaThrLeuGluSerProThrS 173
|||
65 GAAAGCGCTGAACGACACACCGCTGCAGGTCACTGTGACCCAGCGAATG 16

173 eHisPheLeu 176
:: |||||
15 CCGCCTCTCTG 5

seq_name: /cgn2_6/prodata/2/ina/6A_CONB.seq:US-08-526-840B-8

seq_documentation_block:
Sequence 8, Application US/085268403
Patent No. 6001564
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
APPLICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202-4497

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526,840B
FILING DATE: 11-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/304,732
FILING DATE: 12-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586.90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Klebsiella pneumoniae
US-08-526-840B-8

Alignment_scores:
Quality: 113.50 Length: 87
Ratio: 2.027 Gaps: 3
Percent Similarity: 64.368 Percent Identity: 34.483

Alignment_block:
US-09-824-567-2 x US-08-526-840B-8/rev ..

Align seq 1/1 to reverse of: US-08-526-840B-8 from: 1 to: 238

92 GlyLeuThrTyrThrPhelysLeuLysSerAlaPheTrpSerAsnGlyAs 108
|||||
238 GGCTGACC.....TGGTCTGACGGGCAC 216

108 PProLeuThrAlaGluAspPheIleGluSerTrpLysGlnValAlaThrG 125
|||||
215 GGCCATTACAGCGCAGATGCTGTGAGCTGGCAACGGCTGCTCGGC 166

125 InGluValSerGlyIleTyrAla...PheAlaLeuAsn...ProIleLys 139
|||||
165 CGGCACACACCTCCCTACGCCAGCTACCCGGCAATATGCAATATCGCC 116

140 AsnValAlaGlySIIeGlnGluGlyHisLeuSerIleAspHisPheGlyVa 156
|||||
115 AACGCCCGGAGATTGCCCTCGGGCAAAAGGAGACCCGAGAGCTGGGGGT 66

156 IHisSerProAsnGluSerThrLeuValAlaThrLeuGluSerProThrS 173
|||
65 GAAAGCGCTGAACGACACACCGCTGCAGGTCACTGTGACCCAGCGAATG 16

173 eHisPheLeu 176
:: |||||
15 CCGCCTCTCTG 5

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2002, 04:38:12 ; Search time 37.11 Seconds  
(without alignments)  
555,074 Million cell updates/sec

Title: US-09-824-567-2  
Perfect score: 2739  
Sequence: 1 MKRISVGICITILLISVTL.....LSNLSVSTGVNDFRIAKEN 532

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	% length	ID	Description
1	614.5	22.4	545	1	OPPA_BACSU
2	489.5	17.9	543	1	DPPE_BACSU
3	438.5	16.0	537	1	MPPA_ECOLI
4	415	15.2	543	1	OPPA_ECOLI
5	410	15.0	543	1	OPPA_SALT
6	397.5	14.5	535	1	YGIS_ECOLI
7	370	13.5	541	1	OPPA_ECOLI
8	354.5	12.9	514	1	Y213_HABIN
9	337	12.3	652	1	ALIB_STRPN
10	301	11.0	524	1	NKA_ECOLI
11	291.5	10.6	547	1	HBA_HABIN
12	287.5	10.5	516	1	YDGS_ECOLI
13	283.5	10.4	535	1	DPPE_ECOLI
14	275.5	10.1	659	1	AMIA_STRPN
15	272.5	9.9	660	1	ALIA_STRPN
16	249	9.1	512	1	YLIB_ECOLI
17	245	8.9	549	1	SAPA_SALT
18	242.5	8.9	543	1	APPA_BACSU
19	238	8.7	565	1	SAPA_HABIN
20	212	7.7	547	1	SAPA_ECOLI
21	191.5	6.6	600	1	OPPA_ECOLI
22	165.5	6.0	600	1	OPPA_LACIC
23	164	6.0	542	1	XP55_STRPI
24	164	6.0	591	1	YCB0_MYTU
25	156.5	5.7	693	1	AGPA_RHME
26	153.5	5.6	531	1	Y4TO_RHISN
27	148	5.4	663	1	Y4WM_RHISN
28	143	5.2	604	1	YELA_ECOLI
29	141.5	5.2	744	1	HXCL_HABIN
30	126.5	4.6	1432	1	SK13_YRST
31	120.5	4.4	3726	1	ABRI_MOUSE
32	114.5	4.2	412	1	HOFO_ECOLI
33	111	4.1	566	1	YBAE_ECOLI

34	110	4.0	1807	1	YTA2_XENLA
35	109	4.0	441	1	HS82_ASFV
36	108.5	4.0	1237	1	YDY2_SCHPO
37	107	3.9	623	1	RECG_HELPJ
38	107	3.9	701	1	HS90_PODAN
39	107	3.9	1089	1	NMD2_YRST
40	106.5	3.9	2607	1	BACB_BACLI
41	106	3.9	1161	1	FOL_SFV1
42	104.5	3.8	557	1	INRI_HUMAN
43	103.5	3.8	4639	1	DYHC_DROME
44	103	3.8	436	1	ZPI_RAT
45	103	3.8	580	1	MUTL_CHLPPN

## ALIGNMENTS

```

RESULT 1
ID OPPA_BACSU
AC P24141; P23399; STANDARD: PRT; 545 AA.
DE 01-MAR-1992 (rel. 21, Created)
DE 16-OCT-2001 (rel. 40, Last sequence update)
DE 16-OCT-2001 (rel. 40, Last annotation update)
DE Oligopeptide-binding protein oppa precursor.
GN OPPA OR SPOOKA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
CX NBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=91194547; PubMed=1901616;
RA Perego M., Higgins C.F., Pearce S.R., Gallagher M.P., Hoch J.A.;
RT "The oligopeptide transport system of Bacillus subtilis plays a role
RL in the initiation of sporulation."
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=91139580; PubMed=1899858;
RA Rudner D.Z., Ledoux J.R., Iretton K., Grossman A.D.;
RT "The spo0K locus of Bacillus subtilis is homologous to the
RT oligopeptide permease locus and is required for sporulation and
RT competence."
RN [3]
RP Bacteriol. 173:1388-1398(1991).
CC -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE OLIGOPEPTIDE
CC PERMEASE, A BINDING PROTEIN-DEPENDENT TRANSPORT SYSTEM, IT BINDS
CC PEPTIDES UP TO FIVE AMINO ACIDS LONG WITH HIGH AFFINITY. ALSO
CC REQUIRED FOR SPORULATION AND COMPETENCE.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Probable).
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC PROTEIN FAMILY 5.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC
CC EMBL: X56347; CA39787.1; -
CC EMBL: M57689; AAA62687.1; -
CC EMBL: Z99110; CAB1500.1; -
CC PIR: S15230; S15230.
CC PIR: A38447; A38447.
CC HSSP: P06202; 1B52.
CC Subtilist; BG10771; oppa.
CC Interpro: IPR000914; SBP_bac_5.
CC Pfam: PF00496; SBP_bac_5; 1.

```

DR PROSITE: P500013; PROKAR\_LIPOPROTEIN; 1.  
 DR PROSITE: P501040; SBP\_BACTERIAL\_5; 1.  
 KW Peptide transport; Transport; Membrane; Signal; Sporulation;  
 Lipoprotein; Complete proteome.  
 FT SIGNAL 1 20  
 FT CHAIN 1 545  
 FT LIPID 21 21  
 FT CONFLICT 26 26  
 FT CONFLICT 195 195  
 FT CONFLICT 340 340  
 SO SEQUENCE 545 AA; 61524 MW; 8B4B69DFDC4B4B6 CRC64;

Query Match 22.4%; Score 614.5; DB 1; Length 545;  
 Best Local Similarity 29.3%; Pred. No. 8e-36;  
 Matches 158; Conservative 108; Mismatches 249; Indels 25; Gaps 13;

10 IITLSLVYLOGC-----KSSSSSTSGKELAINRDEPNSLDPROVRLSEISLV 61  
 8 VITMLLFTLVLSACGFGGTGSGNGEKKDKGKTTINIKTEPFLSHPLGANDSVSGYI 67  
 62 KATIEGLVQENNLSGNIPEALAEYSLSDGLTYFTKLKSAF-WSNGDPLTADPTEISWK 120  
 68 RQFEELTR-IMADSEPEEGMASKIEISKDGKTYFTTRIDGKKSNDPVTADPFEYAWK 126  
 121 QVATQVSGIYAFALNPYKIVKIQEGHSLIDHFGVHSPNESTLVLTLESPTSHFLKLLA 180  
 127 WALDPNNSQVAYQLYYIKGAEAANTGKSLDVAKAVNDKTLKVELNPTPYETELTA 186  
 181 LPVFFVHKSORTLOSKSLPIASGAFPKY--IKOKW-----IKLSKNHYVNOQVET 233  
 187 FYTYMPLNE-KIAEKNNKMWNTNADDDVNSNGPKMTAMKHSQSTTLKKNQYWDKDKVL 245  
 234 KTTTHIFIDANTAAKLFNOCKLWMOGPPWGERIQEFTLSLQSKGHLHSDFAVGSWLT 293  
 246 KXIDVYMINNNTTELKRFQAGELDMAGMPLQG-LPTESLPLTKDGLHNPYLAGYVYWK 304  
 294 ENINKEPLNNMKLRALASALDKALVSTIFIGRAKTAHDHLLPTNIHSYEHOKQEMAOR 353  
 305 FNTAEKPLDNNVIRKALYSLDRQSTYKVNTOGEOPMMAVPPPTMGPFEDNKEGYFKN 364  
 354 QA-YAKKLFKEALELQIT-AKDLHNLIFPVSSASSLLVOLIROMKESIGFAIPYV 411  
 365 DVYTAKEYLEKLEKEMGSKASDLPKIKLSYN-TDDAIAKAAQAVQEMKKNLGVVEDLD 423  
 412 GKFEALLQADSSGNSLATGWFADPAFALTIFVYPS-GVPPYAINHKDFLEILON 470  
 424 NSEMNYYIDKLSQDYQIGRMGWLGDNDPINFLELFRDXKGNNDGWNPEFKKLNQ 483  
 471 IEEDOHKRSRLVQA-SLYLEFTHIEPTIHDAFORAMKKLSNLGVSPTGVDFRYA 529  
 484 SQTETDKTRAELEKAGIFIDEMP-VAPITYTDTWVDENLKVIMEGTGEVYFRNA 542

RESULT 2  
 DPEPE\_BACSU STANDARD; PRT; 543 AA.  
 AC P26906; O34801;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 30-MAY-2000 (Rel. 39, last sequence update)  
 DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Dipeptide-binding protein dppf precursor.  
 GN dppf OR NCBI:  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 OX NCBI\_TaxID=1423;  
 RN [1]  
 RP SOURCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=92114768; PubMed=1766370;  
 RA Mathopoulos C., Mueller J.P., Slack F.J., Murphy C.G., Patankar S., Bukusoglu G., Sonenshein A.L.;

RT "A Bacillus subtilis dipeptide transport system expressed early  
 RT during sporulation."  
 RL Mol. Microbiol. 5:1903-1913(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Devine K.M.;  
 RT "Sequence of the Bacillus subtilis genome between xlyA and ykoR."  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 CC FOR DIPEPTIDES. PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM  
 CC WITH HIGH AFFINITY. IS EXPRESSED TO FACILITATE ADAPTATION TO  
 CC NUTRIENT DEFICIENCY CONDITIONS, WHICH ALSO INDUCE SPOULATION.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor  
 CC (probable)  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING SPOULATION.  
 CC -1- INDUCTION: NUTRIENT DEFICIENCY CONDITIONS.  
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING  
 CC PROTEIN FAMILY 5.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X56678; CAA0006.1; -  
 DR EMBL; AJ002571; CAA05576.1; ALT\_INIT.  
 DR EMBL; Z99110; CAB1315.1; ALT\_INIT.  
 DR PIR; S16651; S16651.  
 DR HSP; P06202; 1B52.  
 DR Subtilist; BG10846; dppf.  
 DR InterPro; IPR000914; SBP\_bac-5.  
 DR Pfam; PF00496; SBP\_bac-5; 1.  
 DR PROSITE; P500013; PROKAR\_LIPOPROTEIN; 1.  
 DR PROSITE; P501040; SBP\_BACTERIAL\_5; 1.  
 KW Peptide transport; Transport; Membrane; Signal; Sporulation;  
 KW Lipoprotein; Complete proteome.  
 FT SIGNAL 1 22  
 FT CHAIN 23 543  
 FT LIPID 23 23  
 FT CONFLICT 339 339  
 FT SEQUENCE 543 AA; 61822 MW; 723A346197413356 CRC64;

Query Match 17.9%; Score 489.5; DB 1; Length 543;  
 Best Local Similarity 26.0%; Pred. No. 4.9e-27;  
 Matches 143; Conservative 116; Mismatches 244; Indels 47; Gaps 13;

12 IITLSLVYLOGC-----KSSH-SSTSGE--LAINRDEPNSLDPROVRLSEISLV 60  
 12 LALGSLFALMGCTANDQAKRGSHDKAKTGEVYLVNNEPSTFDPIC--FNNNSWO 69  
 61 -VKHIEGLVQENNLSGNIPEALAEYSLSDGLTYFTKLK-SAFWSNGDPLTADPTEISLV 118  
 70 PLNNIMEGLTRIGK-DHEPEPAMAEKMSYKDKKTYFTIRNAKNTMGDVTAGDEFYA 128  
 119 WKQVATQVSGIYAFALNPYKIVKIQEGHSLIDHFGVHSPNESTLVLTLESPTSHFLK 178  
 129 WKRMIDPKKGASSAFIYFIEGGEAYNSGKKMDVYKATAKDFLTLELVADQKFLSV 188  
 179 LALPFFVPHK-----SORTLOSKSLPIASGAFPKYKINIKQKWLKLSKNHYVNOQVET 233  
 189 VSNDAIFPVNEVYKONKMPAESDITVGNGPKYLTETKHDSITMKSQDITMDKDTYKL 248  
 234 KTTTHIFIDANTAAKLFNOCKLWMOGPPWGERIQEFTLSLQSKGHLHSDFAVGSWLT 293  
 249 DKYKMWAMVSDRNTDYQMGQSELD-----TAVPAELSDQLLDQDNVNTVDAAGLYFPR 302  
 294 ENINKEPLNNMKLRALASALDKALVSTIFIGRAKTAHDHLLPTNIHSYEHOKQEMA-- 351

Db 303 FVNWEPQENIRKAFAMAVDQEEIVKYVTKNNKETA-HAFVSPGTPQDCKDREAGG 361  
 QY 352 ----QROAVAKLFEKALELOITAKDLEHLNLFPPVSSKAS-----LLVOLLREQKE 402  
 Db 362 DLKRNESKAKOLEKGMKE-----ENYNKLPATLLIYSTRPKHKAIAIQKILKN 413  
 QY 403 SLGFAIPVIGKEPALLQADLSSGNSFLATGCGWFDPPADPAFLITAIYSSGPVAINHK 462  
 Db 414 SLGVAVYKLANMENWVPLEDQKALKRFQSSSFLPDYADPISLEAFQTSNMTGAMANK 473  
 QY 463 DPLELTONIEODHQRSELSVQSALYLETFPHIIEFYHADAFQAMKRLSLNLGVSPTG 522  
 Db 474 EYDQILIKAKNEADEKTRSLMHQABELLINAPILIVFYVYQVHQLQNVQVIGVIRHVG 533  
 QY 523 VYDFYAKEN 532  
 Db 534 VYDKWADKN 543

RESULT 3  
 ID MPPA\_ECOLI STANDARD; PRT: 537 AA.  
 AC P77348; 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Periplasmic murein peptide-binding protein precursor.  
 GN MPPA OR B1329.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBL\_TaxID:562;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-32.  
 RC STRAIN-K12 / AT980;  
 RX MEDLINE=98155149; PubMed=9495761;  
 RA Pak J.T., Raychaudhuri D., Li H., Normark S., Mengin-Lecreulx D.;  
 RT "Mppa, a periplasmic binding protein essential for import of the  
 bacterial cell wall peptide L-alanyl-gamma-D-glutamyl-meso-  
 diaminopimelate.";  
 RT J. Bacteriol. 180:1215-1223(1998).  
 RL [2]  
 RN STRAIN-K12 FROM N.A.  
 RC MEDLINE=9742617; PubMed=9278503;  
 RA Balthier F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.,  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RT Science 277:1453-1474(1997).  
 RL [3]  
 RN SSOURCE FROM N.A.  
 RP STRAIN-K12;  
 RC MEDLINE=97251357; PubMed=9097039;  
 RA Alba H., Baba T., Fujita K., Hayashi K., Iada T., Isono K., Itoh T.,  
 RA Kasai H., Kashiimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakase S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,  
 RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,  
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.,  
 RT "A 570-bp DNA sequence of the Escherichia coli K-12 genome  
 corresponding to the 28.0-40.1 min region on the linkage map.";  
 RT DNA Res. 3:363-377(1996).  
 RL [4]  
 RN FUNCTION: ESSENTIAL FOR THE UPTAKE OF THE MUREIN PEPTIDE L-ALANYL-  
 GAMMA-D-GLUTAMYL-MESO-DIAMINOPIMELATE. ALSO TRANSPORTS SOME ALPHA-  
 LINKED PEPTIDES SUCH AS PRO-PHE-GLYS WITH LOW AFFINITY. THE  
 TRANSPORT IS EFFECTED BY THE OLIGOPEPTIDE PERMEASE SYSTEM.  
 CC -1- SUBCELLULAR LOCATION: Periplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING  
 PROTEIN FAMILY 5.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC  
 DR EMBL: U88242; AAC38216.1; ALT\_INIT.  
 DR EMBL: AE000231; AAC74411.1; ALT\_INIT.  
 DR EMBL: D90772; BA014932.1; ALT\_INIT.  
 DR EMBL: D90771; BA014922.1; ALT\_INIT.  
 DR HSSP: P06202; IJEV.  
 DR SWISS-2DPAGE: P77348; COLI.  
 DR EcoGene: EG13376; mppa.  
 DR InterPro: IPR000914; SBP\_bac.5.  
 DR Pfam: PF00496; SBP\_bac.5; 1.  
 DR PROSITE: PS01040; SBP\_BACTERIAL\_5; 1.  
 KW Peptide transport; Transport; Periplasmic; Signal; Complete proteome.  
 FT SIGNAL 1 22  
 FT CHAIN 23 537  
 FT PROTEIN PERIPLASMIC MUREIN PEPTIDE-BINDING  
 SO SEQUENCE 537 AA; 59900 MW; C6A17656836DC3AC CRC64;

Query Match 16.0%; Score 438.5; DB 1; Length 537;  
 Best Local Similarity 26.0%; Pred. No. 1.9e-23;  
 Matches 134; Conservative 100; Mismatches 248; Indels 33; Gaps 12;

QY 4 ISVGICTITLLISLVYLOGCKSSHSSTSRGELAINIRDEPRSLDPROYRLSEISLYKH 63  
 Db 5 VAVTCALVSSISLSYAAEVSGTVLAEKQELVRLHKDEPASPDPKAVGLPELQVIRD 64  
 QY 64 IYEGVJOENNISGIPALAEYDLSLSDGLYTFKRL-SAFWSNGDPLAEDFISMKOV 122  
 Db 65 LEFGIYVKNQ-KGEIVGVATQWK-SNDNRIMTPTLBDNAKMAWDGPPVYADQFVYSWQRL 122  
 QY 123 ATQEVSGTYAF-ALNDIKNRKIQEGHSLDHFVSPNESTLYVTLSPSHLKTLL 180  
 Db 123 VDPKTLSPFAFAPALAGINNAQAIIIDKATPDQGLVAVNAHLKIQDPLPMFVNILTA 182  
 QY 181 LPVFPPYHK-----SQTLOSKSLPTASGATYPKNIKOKIKLSKNPHYNOGYEYRTI 236  
 Db 183 NFAPFPQKANVESGKEMTKPGLNIGNAVYKERYNEKLVVYPVTHYMDNAKTIVLQKV 242  
 QY 237 TIRFTPDNATLAFLFNOGKLNMGPPGGERIPDETLSNLSQK-----GHLSHSDVAGTS 290  
 Db 243 TELPINESAATRYIAGDID-----TTSEFPKNMYOKLIDIFQGVYTPPOLGTY 293  
 QY 291 WLTNINIKFPLNMLKREALASALDKEALVSTIFLGAKTADHLPTNINHY-PEHOK-Q 348  
 Db 294 YVANTQKGPPTADQVRRLASWTIDRLMTKVLGGEKPAHMFPPDVTAGTPEPPEPFE 353  
 QY 349 EMAQRO-AVAKKLFKEALEEELQITAKDLEHLNLFPPVSSASSLLVOLLREOKESLGF 406  
 Db 354 QMSQEEILMAQKILLASAGYQPKL---LTLVNTSENQKTAI-AVASMKNLGLV 408  
 QY 407 AIPVIGKEPALLQADLSSGNSFLATGCGWFDPPADPAFLITAIYSSGPVAINHKFELE 466  
 Db 409 DVKLQNGEMKTYIDSNTNGNDIVIRASVGDVNEPFTLTLTSHSGNISRFNNPAYDK 468  
 QY 467 ILQNIREDQDHQRSELSVQSALYLETFPHIIEFY 501  
 Db 469 VLAQASTENTYKARNADYNAEKILMEQAPIARIY 503

RESULT 4  
 ID MPPA\_ECOLI STANDARD; PRT: 543 AA.  
 AC P23843; P76829;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Periplasmic oligopeptide-binding protein precursor.





QY 64 IYEGVLQVNNISGNIPALAEYSISDGLTYTFKLK-SAFWSNGDPLTAEDFIESKQV 122  
 Db 69 LFEGLL-ISDVEGHSPGVAEKWE-NKDFKVTWFLHRENAKSDGTPVTAHDFVYSWQRL 126  
 QY 123 ATQEVSGIYAFAL-NPIKNVRKIQGHLSIDHFGVHSPNESILVLTLESPTSHFLKLLA 180  
 Db 127 ADPNTASPYASYLOYGHIANIDDIAGKKPATDILGVKALDDHTFEVTLSEPVYFYKLLV 186  
 QY 181 LPVFPFVHKS-----QRTLSQSLPIASGAFYKPIKQKWLKSKNPHYINQSOVETK 234  
 Db 187 HPSVSPVPSKSAVEKFGDKWTQANI-VINGAYKLKNVNVNERIVLERNQYWDNAKTVIN 245  
 QY 235 TITHTFIDPANTAAKLFNOKLWNOGPEWGERIPQETLSLQSK--GHLHSDVAGTSWL 292  
 Db 246 QVTYLPISSEVTDVNYRSGEIDMT-----YNNMPIELFOKLKEIPNEVRVDPYLCTYY 301  
 QY 293 TFINKEPLNNMKLREALASALDEALYSTIFLGRAKTADHLLPTNIHSY----- 342  
 Db 302 EINNOKAPFNDVRVTRALKLALDRDIIVNKV---KNQGD--LPA--XYSTPPVTDGAKL 353  
 QY 343 --PEHQKQMAORQAVAKKLFKEALEEQITAKDLEHLNLIFPVSSASSLLVOLIREQW 400  
 Db 354 VEPEWFKWSQKRNEEAKKLAEEA---GFTADKPLTFDOLLNTSDLHKKLAIAVASIW 408  
 QY 401 KESTGFAIPVIGKEFALLOADSSNFSLATGWPADFADPMAPLTIFFAYPSGVPPYAIN 460  
 Db 409 KKNLGVNVNLENOEMKFTLDRHOQTFDVARAGWCADYNEPT'SFLNTMLSDSSNNTAHYK 468  
 QY 461 HKDFLEIQNIEQEDHOKRSELVSQASLYLETFFHIEPIYH 502  
 Db 469 SPAPDKLIADTLKVADDTQORSELYAKAEQQLDKDSALVPPVY 510  
 RESULT 6  
 ID YGIS\_ECOLI STANDARD; PRT; 535 AA.  
 AC Q46863;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Putative binding protein *ygis* precursor.  
 GN YGIS OR B3020.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=X12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [2]  
 RP IDENTIFICATION BY MASS SPECTROMETRY.  
 RX MEDLINE=99420866; PubMed=10493123;  
 RA Pountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;  
 RT "Enrichment of low abundance proteins of Escherichia coli by  
 RT hydroxyapatite chromatography.";  
 RL Electrophoresis 20:2181-2195(1999).  
 CC -!- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT  
 CC SYSTEM.  
 CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING  
 CC PROTEIN FAMILY 5.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U28377; AAA69188.1; -;  
 DR EMBL; AE000384; AAC76056.1; -;  
 DR HSSP; P06202; 1B52.  
 DR EcoGene; EGI3021; *ygis*.  
 DR InterPro; IPR000914; SBP\_bac\_5.  
 DR Pfam; PF00496; SBP\_bac\_5; 1.  
 DR PROSITE; PS01040; SBP\_BACTERIAL\_5; FALSE\_NEG.  
 KW Transport; Periplasmic; Signal; Complete proteome.  
 FT SIGNAL 1 20 POTENTIAL.  
 FT CHAIN 21 535 PUTATIVE BINDING PROTEIN YGIS.  
 SQ SEQUENCE 535 AA; 60694 MW; B0F56723071A54EF CRC64;  
 Query Match 14.5%; Score 397.5; DB 1; Length 535;  
 Best Local Similarity 25.8%; Pred. No. 1.4e-20;  
 Matches 126; Conservative 82; Mismatches 233; Indels 47; Gaps 12;  
 QY 39 NIRDEPRSLDPRQVRLSEISLVKHYEGLVQENNLGNIEPALAEYDLSLSDGLTYTFK 98  
 Db 38 NNHSDPGTLDPKQVEENTAAQIVLDLFEGLVMDG-EGQVQPAQAEWEILDGGRKRYIF 96  
 QY 99 LKSAF-WNSGDPLTAEDFTESKQVATQEVSGIYA--FALNPIKNVRKIQGHLSIDHFG 155  
 Db 97 LRSLGQSDGQPLTAEDFVLGMQRAVDPKTASPFAGYLAQAHINNAATVACKADVTSLG 156  
 QY 156 VHSNPESLVVLTLESPTSHFLKILALPVFFPV--HKSQRTLSQSLP---IASGAFYPKN 210  
 Db 157 VKATDDRTLEVTLEQVPWFMTMLAWPTLFPVPHVIAKHGDSWSPENWVYNGAFVLDQ 216  
 QY 211 IKQKWIKLSKNPHYINQSOVETKITITHTFIDPANTAALKFNQK--LNWQGPMPGERIP 268  
 Db 217 WYVNEKITARKNEKYRDAQHTVLQOQVEYALDLSNVTCYNYRAGEVDLTW-----VP 268  
 QY 269 QETLSNLQSK--GHLHSDVAGTSMLTININKFPLNNMKLREALASADKALYSTIFLG 326  
 Db 269 AQQIPAEKSLPGLRLTIIPRLNSEYINLEKPPFNDVRVRALYLTDRQ-LIAQKVLG 327  
 QY 327 RAKTADHLLPTNIHSYP-----EHQKQMAORQAVAKKLFKEA-----LEELQITAKD 374  
 Db 328 LRTPATLTTPPEYKGFSAITFDLOKPNSEVAMAKALLKQAGYDASHPLRFELFYNNYD 387  
 QY 375 LEHLNLIFPVSSASSLLVOLIREQWESLGFATPIVIGKEFALLQADLSGNSFLATGGW 434  
 Db 388 LHKEITAI-----ALSEWKKWLGAQVTLRTMEWKTLYLDARRAGDFMLSRQSW 434  
 QY 435 FADFADPMAPLTIFFAYPSGVPPYAINHKDFLEILQNTIEQEDHOKRSELVSQASLYLET 494  
 Db 435 DATYNDASSFLNTLKSDEENVGHWKNAQYDALLNQATQITDATKRNALYQQAEEVINQ 494  
 QY 495 HITEPIYH 502  
 Db 495 APLTIPIY 502  
 RESULT 7  
 ID OPBA\_HAEIN STANDARD; PRT; 541 AA.  
 AC P71370;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Periplasmic oligopeptide-binding protein precursor.  
 GN OPBA OR H1124.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 CC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.



```
QY   395 LIREQKESLGFAIPVKGFEALLQADLSSGNFSLATGWFADFADPWAFLLTFIFAYPSGV 454  
    | : ||| :  
Db   416 LENQEKKTYID-----SRRAGRYDYVARGWADYNQAQTFFNGVFNLSSN 460  
  
QY   455 PPVAINHKKDFLEILLNQTEODHQKRSELVSQASLYLTFTHTIETPH 502  
    | : ||| :  
Db   461 NTAKYANPYDKMAEYSAAATADEGRAKAYAARAEILKDYGIVPIFN 508  
  
RESULT      8  
Y213_HAEBIN STANDARD; PRT; 514 AA.  
AC R44572;  
DT 01-NOV-1995 (Rel. 32, Created)  
DI 01-NOV-1995 (Rel. 32, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DR Putative binding protein HI0213 precursor.  
GS HI0213.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OX Haemophilus.  
OX NCBI_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RD / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K.C., Sutton G., Fitzhugh W., Fields C.A., GCayne J.D.,  
RA Scott J.P., Shirley R., Liu L.I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Utterback I.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchham J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Ghnrm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
RT influenzae Rd".  
RL Science 269:498-512(1995).  
CC -! FUNCTION: PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM  
CC (BY SIMILARITY).  
CC -! SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor  
CC (Probable).  
CC -! SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING  
CC PROTEIN FAMILY 5.  
CC -----  
CC This SWISS-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to licens@isb-sib.ch).  
CC -----  
CC EMBL; U32706; AAC21881.1; -.  
CDR DR HSP; P06202; LB52.  
DDR DR TIGR; HI0213; .  
DDR DD InterPro; IPRO00914; SBP_bac_5.  
DDR DD Pfam; PF00496; SBP_bac_5; 1.  
DDR DD PROSITE; PS00013; PKOKAL_LIPOPROTEIN; 1.  
DDR DD PROSITE; PS01040; SBP_BACTERIAL_5; 1.  
KW Hypothetical protein; TransposT; Membrane; Signal; Lipoprotein;  
complete proteome.  
XX SIGNAL 1 23 POTENTIAL.  
XX CHAIN 24 514 PUTATIVE BINDING PROTEIN HI0213.  
XX FT 24 24 N'-ACYL DIGLYCERIDE (PROBABLE).  
XX LPID 24 24  
SQ SEQUENCE 514 AA; 58876 MW; 789186CA328DBDEC CRC64;
```

Query Match                12.9%; Score 354.5; DB 1; Length 514;  
Best Local Similarity     24.2%; Pred.No.1.4e-17;  
Matches 128; Conservative 105; Mismatches 213; Indels 83; Gaps 20;



Db 460 VLGSNDNVI--DIQQLTSDFSSGYFAQTAQKDYLDYHGKNGPDYQDPSTYLDIENFN 517  
 QY 452 S-----GVPVAINHKD-----FLEILONTEBOHQKRSSELVSOASLYLETHII 497  
 Db 518 SGGFLQNLGLEPEANDKAKAVGLDVDTOMLEANKEDQPAKRYEKYADIQAWLIDSSLV 577  
 QY 498 EPIVHDAFOFAMNKKL---SMLGVSP--GVVDFERYAK 530  
 Db 578 LPSVSRGGTSLRRVTPFAAAYGLTGKGVESYKYLK 614

RESULT 10  
 NIKAI\_ECOLI  
 ID NIKAI\_ECOLI STANDARD; PRT; 524 AA.  
 AC P33590;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Nickel-binding periplasmic protein precursor.  
 GN NIKAI OR B3476.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=95020649; PubMed=7934931;  
 RA Navarro C., Wu L.-F., Mandrand-Berthelot M.-A.;  
 RT "The nik operon of Escherichia coli encodes a periplasmic binding-  
 protein-dependent transport system for nickel.";  
 RL Mol. Microbiol. 9:1181-1191(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=94316500; PubMed=8041620;  
 RA Sofia H.J., Burland V., Daniels B.L., Plunkett G. III, Blattner F.R.;  
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the  
 region from 76.0 to 81.5 minutes.";  
 RL Nucleic Acids Res. 22:2576-2586(1994).  
 CC -!- REPRESENTS THE NICKEL BINDER.  
 CC -!- SUBCELLULAR LOCATION: Periplasmic (Probable).  
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING  
 PROTEIN FAMILY 5.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X73143; CA851659.1; .  
 CC EMBL; U00039; AAB18451.1; .  
 CC EMBL; AE000423; AAC76501.1; .  
 CC PIR; S39594; S39594.  
 CC EcoGene; EGI2075; nika.  
 CC InterPro; IPR000914; SBP\_bac\_5.  
 CC Pfam; PF00496; SBP\_bac\_5; 1.  
 CC PROSITE; PS01040; SBP\_BACTERIAL\_5; 1.  
 CC TRANSPORT; Nickel; Signal; Periplasmic; Complete proteome.  
 FT SIGNAL 1 22  
 FT CHAIN 23 524 NICKEL-BINDING PERIPLASMIC PROTEIN.  
 SQ SEQUENCE 524 AA; 58719 MW; CB2E3C1CDBCE42396 CRC64;

Query Match 11.0%; Score 301; DB 1; Length 524;  
 Best Local Similarity 24.2%; Pred. No. 8.4e-14;  
 Matches 119; Conservative 93; Mismatches 209; Indels 70; Gaps 18;  
 QY 64 IVEGLVQNNLSGNIPEALAEYLSLSDGLTYTFKLK-SAFWSNGDPLTAEDFIESNKKV 122

Db 55 VVEPLV-KYQADGSVFWLAKSWTHSEDKTWTFTLRDDVAFSGNPFDAEAAENFRAV 113  
 QY 123 ATQEVSGIYAFALNPINVKRIQEGHLSIDHFGVHSPNESTLVLTLESPTSHFLKLLALP 182  
 Db 114 LDNRORHAWLELANQIVDVKAL-----SKTELQITLKSAYVFFLOELALP 158  
 QY 183 VFF-----PVHKSORTLOSXSPLASGAFYPKNIKOKQWIKLSKNPHYTNOSQVETKIT 237  
 Db 159 RPFRIAPSFKNHETMNGIKPIGTGPWILQESKLNQYDVEVRNENYNGKPA-IKKIT 217  
 QY 238 IHFIPDANTAALFNOGKLNWQPPWGER--IPQETLSNLSQSKGHLHS--FDVAQTSWLT 293  
 Db 218 FNVIPDPTTRAVAFETGDIIDL---YMGEGILLDTFARFSONPAYHTOLSQPIETVMLA 274  
 QY 294 FNINKFPLNNKMLREALASALDKALVSTIFLGRAKTADHLLPTNI-----HSYPEH 345  
 Db 275 LNTAKAPTNELAVREALNVAVNKKSLIDNALYGTQQVADTIFAPSPVYANLGLKPSOYDP 334  
 QY 346 QKQEMARQAYAKKLFKEALELQITAKDLEHLN-----LIFFVSSSASSLLVLQIRE 398  
 Db 335 QK-----AKALLEKAGWTLF-AGKDIREXNGQPLRIELSTFGTDALSKSMAEIIQA 384  
 QY 399 QWKESLGAIPVIGKEFALLQADLSSGNFSLA-TGGWFADFPADPMFLTIFAYPS----- 452  
 Db 385 DMQ-IGADVSLIGEEESIYARQDRGFWIFHTWAGPY-DPHAFSSMRVPSHADFQ 442  
 QY 453 ---GVPPYAINKKHLEILQNIQEQDQKRSSELVSOASLYLETHIETPIYHDAFOFAM 509  
 Db 443 AQOGLADKPLIDKEIGEVLAETHDTQQUALYRLTLR--LHDEAVYL--PISYISMMVVS 498  
 QY 510 NKXLSNLGVSP 520  
 Db 499 KPELGNIPYAP 509

RESULT 11  
 HBPA\_HAEIN  
 ID HBPA\_HAEIN STANDARD; PRT; 547 AA.  
 AC P33950;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Heme-binding protein A precursor (Hemin-binding lipoprotein).  
 GN HBPA OR DPPA OR HI0853.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OX NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN=DL42 / SEROTYPE B;  
 RX MEDLINE=92267636; PubMed=1339409;  
 RA Hanson M.S., Slaughter C., Hansen E.J.;  
 RT "The hbpA gene of Haemophilus influenzae type b encodes a  
 heme-binding lipoprotein conserved among heme-dependent Haemophilus  
 species.";  
 RL Infect. Immun. 60:2257-2266(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RD / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus  
 influenzae Rd.";

RL Science 269:496-512(1995).

RP [3] POSSIBLE FUNCTION, AND SUBCELLULAR LOCATION.

RC STRAIN-DL42 / SEROTYPE B;

RX MEDLINE=91251755; PubMed=2041470;

RT "Molecular cloning, partial purification, and characterization of a

RA hamelin-binding lipoprotein from *Haemophilus influenzae* type b.";

RL Mol. Microbiol. 5:267-278(1991).

CC -!- FUNCTION: IMPORTANT ROLE IN HEME ACQUISITION OR METABOLISM.

CC -!- SUBCELLULAR LOCATION: Attached to the inner membrane by a lipid anchor.

CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING PROTEIN FAMILY 5.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; M88134; AAA73214.1; ALT\_SEQ.

DR EMBL; M84028; AAA24962.1; .

DR EMBL; U32767; AAC22512.1; ALT\_INIT.

DR PIR; A43832; A43832.

DR HSP; P23847; 1DPE.

DR TIGR; H10853; .

DR InterPro; IPR000914; SBP\_bac\_5.

DR Pfam; PF00496; SBP\_bac\_5; 1.

DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.

DR PROSITE; PS01040; SBP\_BACTERIAL\_5; 1.

FW Inner membrane; Signal; Lipoprotein; Complete proteome.

FT SIGNAL 1 18 PROBABLE.

FT CHAIN 19 547 HEME-BINDING PROTEIN A.

FT LIPID 19 19 N-ACYL DIGLYCERIDE (PROBABLE).

FT VARIANT 48 49 KA -> NS (IN STRAIN DL42).

FT VARIANT 181 181 T -> N (IN STRAIN DL42).

FT VARIANT 240 240 H -> N (IN STRAIN DL42).

FT VARIANT 343 343 T -> I (IN STRAIN DL42).

FT VARIANT 375 375 A -> V (IN STRAIN DL42).

SQ SEQUENCE 547 AA; 60660 MW; 7755E2164A40CB31 CRC64;

Query Match 10.6%; Score 291.5; DB 1; Length 547;

Best Local Similarity 23.1%; Pred. No. 4.2e-13;

Matches 128; Conservative 95; Mismatches 259; Indels 73; Gaps 18;

QY 13 LLSLSVLQCKSSHS-----STSRGELAINIRDEPSRLDPRO-VLLSEISLV 61

DB 9 LAATLVLAACDQSSANKSTAQTEAKSSNNVTTCAPLGFSPALITEGTSYNA 68

QY 62 KHIYEGLVQNNLSGNIPEALAEYSLSDDLTYTFKLKSA-----FWSNGDPLTAE 113

DB 69 QQVNRULVEKKGSTDIPEALAESWEISDGLSVTFHLRGVGFHTFTPTD-FNAD 127

QY 114 DFIESWKQVATQEVSGIYAFALNPKVNRKIQEGHLSIDHF-----GVHSPNESTLVTL 168

DB 128 DWVFSFORQLDPN-----HPYHNVSKGTYPFYFKAMKFPPELLKSVKVDNTITIL 178

QY 159 ESPTSHFLKLLALPVPFVHKSQ-----RTLQSKSLPIASGAFYPKNIKQKWI 217

DB 179 NKTDATPLASLGMDFISYSAEYADSLKAGKPEFLDSR--PVGTGPFVFDVYKTDQAI 235

QY 218 KLSKNPHYNQSOVETKTIITHTIPDANTAALFNQGLNW-QGPPWGERIPDET---LS 273

DB 236 QYVAHENYW-KGRTPDLRLVSIIVPDATTRYAKLQAGTCDDLLFPNVDIAKMKRTPKVQ 294

QY 274 NLQSKGHLHSFDVAGTSWLTNTINKFPLNNKRLREALSALDKALYSTIFLGRKATADH 333

DB 295 LLEQKG-----LNVAIFATNEKAFDNNVKVQALNVAVDKKAITEVYQAGTSAXN 347

QY 334 LLPTNIHSYPEHQEMQAQOAVAKKLEALEELQITAKDLHLNLIFFV---SSASS 390

DB 348 PLPPTIWSYND-EIQDYPYDEKAKQLLAE-----GTPNGFETDFIQVIRASNP 401

QY 391 LLVQLIRQWESLGFAIPVIGKEFALLQADLSSGNFSLATGGWFAFDAPMAFLTPAY 450

DB 402 RMAELTMDWAK-IGVKTNPVTYEWADYKRAKEGELTAGIFGSGDGDPDFLSPILG 460

QY 451 PSGV---PPYAINHKDELEILQNIQEQDQHKRSELVSQASLILETHIETIYDAFQF 507

DB 461 SSNIGNSNWFNFSEFDALNEAIGLTNKERAKLYKQAVIVHQAQWIPVAHSVGF 520

QY 508 ANKKLSNLGVSPG 522

DB 521 PLSPRVKGYVQSPG 535

RESULT 12

ID YDVS\_ECOLI STANDARD; PRT; 516 AA.

AC P76128; P77769; P76874;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Putative ABC transporter periplasmic binding protein ydds precursor.

GN YDVS OR B1487.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI\_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;

RT "The complete genome sequence of *Escherichia coli* K-12.";

RL Science 277:1453-1474(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=K12.

RX MEDLINE=97251357; PubMed=9097039;

RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M., Mori K., Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Tagami H., Takeda J., Takeuchi K., Takeuchi Y., Wada C., Yamanoto Y., Horiuchi T.;

RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 28.0-40.1 min region on the linkage map.";

RL DNA Res. 3:363-377(1996).

CC -!- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM YDVS.

CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).

CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING PROTEIN FAMILY 5.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AE000245; AAC74560.1; .

DR EMBL; D90789; BAA15142.1; ALT\_INIT.

DR EMBL; D90790; BAA15152.1; ALT\_INIT.

DR EMBL; D90791; BAA15158.1; ALT\_INIT.

DR EcoGene; EG13790; ydds.

```

DR InterPro; IPR000914; SBP_dac_5.
DR Pfam; PF00496; SBP_dac_5; 1.
DR PROSITE; PS01040; SBP_BACTERIAL_5; 1.
KW Hypothetical protein; Transport; Periplasmic; Signal;
KW Complete proteome.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 516 PUTATIVE ABC TRANSPORTER PERIPLASMIC
FT SEQUENCE 516 AA; 57641 MW; 86E1C32CC3E06FB9 CRC64;
SQ

Query Match          10.5%; Score 287.5; DB 1; Length 516;
Best Local Similarity 22.1%; Pred. No. 7.3e-13;
Matches 123; Conserved 109; Mismatches 208; Indels 117; Gaps 20;

QY 16 LSVVLOQKESHSTSGELAINIRDEPRSLDPROVRLLEISLVKHIVEGLVQ----E 71
DQ 12 LALVLAETNPVAAHAPKDMVLVIGKAAQPTLDPAVTIDNDWTVPVSORLVQKIDG 71
QY 72 NNLGNIIEPALAEYSSDGLATYFKLK-SAFWSNGDPLTAEDFIESWKOVATQEVSGI 130
DQ 72 DKGSTDVEGLDASSWKASDDQKWTFTLKDNAKFAAGTGPVIAEAVKLSFERLL 124
QY 131 YAFALMPIKNVRKIQEG-----HLSIDHFGVHSPNESTLWLTLESTPSHFLKLI 180
DQ 125 -----KIGQGPAAEFKPKLID-----APDEHTVKFTLSQTPAPFLYTLANDG 167
QY 181 -----LPVFFPVHK-----SORTLQSKSLPIASGAFYKNIKQKOWIKLSKNPHYNQ 228
DQ 168 ASIIINPAVLKEHAADDAARGFLAQT-----AGSGPFMLKSKWQKQQLVLPNPHYG- 219
QY 229 SOVETKTIITHIFDANTAALFNQCKLWNOGPPWGERIPQETLSNL--QSRGHLHSFDV 286
DQ 220 NKPNEKRVSVKIGESASRRLQSLRGDID-----IADALPVDQLNALKQENKVNVAEYPS 274
QY 287 AGTSLWTFNFKFPLNNMKLRALASALDKALYSTITFLGRATADHLLPTNIHSYPEHQ 346
DQ 275 LRVTYLVNNSKAPLQAQDLRAISWSTDYOGVNGILSGKQMGPIEGMWGYD--- 331
QY 347 KOEAAQQAQAKKLKEALEEIQITAKOLEHLNLF-----PVSSASSLLVQLIRE 398
DQ 332 -----ATAMQVNHDETRAKAEKVKTSKPTS-LTFLYSDNDPWEPIALATOSSL----- 380
QY 399 QWKSGLGFAIPVIGKEFALLQADLSSGNFSLATGCGFADPADPMAFLTIF--AYPSGVPP 456
DQ 381 ----NKLIIIVKLEKLANIMRDRVKGCDYDIAIGNNSPDPADPYEMMYWESDKGLP- 436
QY 457 YAINHKDFLE-----ILQN-----IEQEDHOKRSELVSQASLYLETHIETPIYHD 503
DQ 437 ---GNRSFVENSEVDKLLRNALATTDQTRDYQQAQKIVIDDAAYVY-----LFQK 486
QY 504 AFOFAMNKLNLGVSP 520
DQ 487 NYQLANKEKVGFEVFP 503

RESULT 13
DPPA_ECOLI
ID DPPA_ECOLI
AC P23847;
DT 01-NOV-1991 (Rel. 20, Created)
DT 16-OCT-2001 (Rel. 20, Last sequence update)
DE Periplasmic dipeptide transport protein precursor (Dipeptide-binding
DE protein) (DBP).
GN DPPA OR B3544..
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;

```

```

RX MEDLINE=91100289; PubMed=1702779;
RA Olson E.R., Danyak D.S., Jurss L.M., Poorman R.A.;
RT "Identification and characterization of dppA, an Escherichia coli
RT gene encoding a periplasmic dipeptide transport protein.";
RL J. Bacteriol. 173:234-244(1991).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-50.
RC STRAIN=K12;
RX MEDLINE=92065799; PubMed=1956284;
RA Abouhamad W.N., Manson M., Gibson M.M., Higgins C.F.;
RT "Peptide transport and chemotaxis in Escherichia coli and Salmonella
RT typhimurium: characterization of the dipeptide permease (Dpp) and the
RT dipeptide-binding protein.";
RL Mol. Microbiol. 5:1035-1047(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MM500;
RX MEDLINE=95231288; PubMed=7536291;
RA Abouhamad W.N., Manson M.D.;
RT "The dipeptide permease of Escherichia coli closely resembles other
RT bacterial transport systems and shows growth-phase-dependent
RT expression.";
RL Mol. Microbiol. 14:1077-1092(1994).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes.";
RL Nucleic Acids Res. 22:2576-2586(1994).
RN [5]
RP SEQUENCE OF 29-40.
RC STRAIN=K12 / BMG2;
RX MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robinson K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12.";
RL Electrophoresis 18:1259-1313(1997).
RN [6]
RP DISULFIDE BONDS.
RA Duntzen P.;
RL Submitted (JUN-1995) to the SWISS-PROT data bank.
RN [7]
RX X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS).
RA Duntzen P., Mowbray S.L.;
RT "Crystal structure of the dipeptide binding protein from Escherichia
RT coli involved in active transport and chemotaxis.";
RL Protein Sci. 4:2327-2334(1995).
RN [8]
RX X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RA Nickitenko A.V., Trakhanov S., Quilocha F.A.;
RT "2-A resolution structure of dppA, a periplasmic dipeptide
RT transport/chemosensory receptor.";
RL Biochemistry 34:16585-16595(1995).
CC -!- FUNCTION: DIPEPTIDE-BINDING PROTEIN OF AN OSMOTIC-SHOCKABLE
CC TRANSPORT SYSTEM. DPPA IS ALSO REQUIRED FOR PEPTIDE CHEMOTAXIS.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC PROTEIN FAMILY 5.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M35045; AAA23707.1; -.
CC EMBL; X58051; CAA41090.1; -.
DR
DR

```

DR EMBL; U08399; AAA23702.1; -  
 DR EMBL; L08039; AAB18522.1; -  
 DR EMBL; AE000431; AAC76569.1; -  
 DR PIR; S15292; S15292.  
 DR PIR; A39194; A39194.  
 DR PDB; 1DPE; 17-AUG-96.  
 DR PDB; 1DPP; 17-AUG-96.  
 DR SWISS-2DPAGE; P23847; COLI.  
 DR ECO2DBASE; G059.9; 6TH EDITION.  
 DR Ecogene; EGI0248; dppA.  
 DR InterPro; IPR000914; SBP\_bac\_5.  
 DR Pfam; PF00496; SBP\_bac\_5; 1.  
 DR PROSITE; PS01040; SBP\_BACTERIAL\_5; 1.  
 KW Peptide transport; transport; periplasmic; signal; Chemotaxis;  
 FT 3D-structure; Complete proteome.  
 FT SIGNAL 1 28  
 FT CHAIN 29 535 PERIPLASMIC DIPEPTIDE TRANSPORT PROTEIN.  
 FT DISULFID 34 262  
 FT DISULFID 450 463  
 SQ SEQUENCE 535 AA; 60293 MW; 3F7C4756E2A8C2F0 CRC64;

Query Match 10.4%; Score 283.5; DB 1; Length 535;  
 Best local similarity 24.6%; Pred. No. 1.5e-12;  
 Matches 122; Conservative 85; Mismatches 223; Indels 65; Gaps 20;

QY 64 IYELVQENLSCNIEPALAEYLSLSDGLTYFKL-KSAFWSNGD-----PLTAEDFI 116  
 Db 60 LYNRLEVFYKIGTEVTEVPLGAKEVSESDGKTYTFHLRKGVKWHDNKEFKETRELNADVV 119  
 QY 117 ESMQVATQVSGYVAFALNPIKVRKIQEGLHSLDHP-GVHSP-----NESTLVV 166  
 Db 120 FSDRQKN-----AQNP---YHKVSGG---SYEYFEGMGLPELISEVKVVDNTVQF 165  
 QY 167 TLESTPTSHFLKLALPVFFVHKSORTLQSKS-----LPTASGAFYPKNKKQKQWIK 218  
 Db 166 VLTREAPFLADLANDFASILSEYADAMKAGTPEKLDLNPITGTFQLOQYQKDSRIR 225  
 QY 219 LSKNPHYV-NGSQVETWITIHPIPDANTAKLFNQK---LWQGPFPWGERIPQETLSN 274  
 Db 226 YKAFQGVWGTQPKQIDTLVFSI---TPDASVRYAKLQKNEQVMPYNPADIARMKQDKSIN 283  
 QY 275 LOSKGLHLSFDVAGTSMLFNINKPPLNNMKLREALASLDKALYSTIFLGRKATADHL 334  
 Db 284 LMEMPGLN-----VGYSYNNQKPLDDVYKQALTYAVNKKDAIKAVYQAGVSAKNL 337  
 QY 335 LPTNHSYPHQKQMAQKQVAKKLFKEALELQITAKLEHLNIPVSVSSASSLLVQ 394  
 Db 338 IPTPMWGYND-DVQDYTYDPERAKALKKEAGLEKGFSI-DLWAMPVQREYNPNARR-MAE 394  
 QY 395 LIRQWKESLGFAIPVIGKEFALLQADLSNGFSLATGWFADFP-MAFLTTF---AY 450  
 Db 395 MIQADWAK-VGVQAKLVTEWGEYKRAKGEHQVTVMGWTGNDGDPNFFATLFSKAAS 453  
 QY 451 PSGVPPVAINKHDFLEILQNIQEODHOKRSELSVQASLYLTFEHLIP---IYHDAQF 507  
 Db 454 EQGSNYSKWCYKPFEDLQIPARATDNDKNRVLYKQAVVM---HDQAPALIIAHSTVFE 510  
 QY 508 AMNKKLSNLGVSPGT 522  
 Db 511 PVREKVGKGYVDPLG 525

RESULT 14  
 ID AMIA\_STRPN  
 AC AMIA\_STRPN STANDARD; PRT; 659 AA.  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 GN Oligopeptide-binding protein amIA precursor.  
 GN AMIA OR SPI891.  
 OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 RX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R800;  
 RX MEDLINE=90279506; PubMed=2352474;  
 RA Allouing G., Trombe M.C., Claverys J.-P.;  
 RT "The ami locus of the Gram-positive bacterium Streptococcus pneumoniae  
 is similar to binding protein-dependent transport operons of gram-  
 RT negative bacteria.";  
 RL Mol. Microbiol. 4:633-644(1990).  
 RN [2]  
 RP REVISIONS.  
 RA Claverys J.-P.;  
 RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TIGR4;  
 RX MEDLINE=21357209; PubMed=11463916;  
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,  
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
 RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,  
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
 RT "Complete genome sequence of a virulent isolate of Streptococcus  
 pneumoniae";  
 RL Science 293:498-506(2001).  
 RN [4]  
 RP SEQUENCE OF 1-28 FROM N.A.  
 RC STRAIN=R800;  
 RX MEDLINE=90060772; PubMed=2684766;  
 RA Martin B., Allouing G., Boucraut C., Claverys J.-P.;  
 RT "The difficulty of cloning Streptococcus pneumoniae mal and ami loci  
 in Escherichia coli: toxicity of malX and amiA gene products.";  
 RL Gene 80:227-238(1989).  
 RN [5]  
 RP PARTIAL SEQUENCE.  
 RX MEDLINE=89091147; PubMed=3208757;  
 RA Gilson E., Allouing G., Schmidt T., Claverys J.-P., Dudler R.,  
 RA Hofnung M.;  
 RT "Evidence for high affinity binding-protein dependent transport  
 systems in Gram-positive bacteria and in Mycoplasma.";  
 RL EMBO J. 7:3971-3974(1988).  
 CC -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM  
 FOR OLIGOPEPTIDES; PROBABLY AN OLIGOPEPTIDE BINDING PROTEIN.  
 CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID  
 ANCHOR.  
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING  
 PROTEIN FAMILY 5.  
 CC -!- CAUTION: THE REVISED SEQUENCE OF AMIA NOW INCLUDES, IN THE  
 C-TERMINAL SECTION, THE SEQUENCE OF AN ORF WHICH WAS PREVIOUSLY  
 KNOWN AS AMTB.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X17337; CAA35213.1; -  
 DR EMBL; AE007479; AAK75962.1; -  
 DR PIR; S11148; S11148.  
 DR PIR; S11149; S11149.  
 DR TIGR; SPI891; -  
 DR InterPro; IPR000914; SBP\_bac\_5.  
 DR Pfam; PF00496; SBP\_bac\_5; 1.  
 DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.



DR PROSITE; PS01040; SBP\_BACTERIAL5; 1.  
 KW Peptide transport; Transport; Signal; Membrane; Lipoprotein;  
 FT Complete proteome.  
 FT SIGNAL 1 22 PROBABLE.  
 FT CHAIN 23 659 OLIGOPEPTIDE-BINDING PROTEIN AMIA.  
 FT LIPID 23 23 N-ACYL DIGLYCERIDE (PROBABLE).  
 FT CONFLICT 23 23 S->R (IN REF. 1).  
 FT CONFLICT 421 421 A->R (IN REF. 1).  
 FT CONFLICT 557 557 R->K (IN REF. 1).  
 FT CONFLICT 562 562 L->F (IN REF. 1).  
 FT CONFLICT 622 622 S->L (IN REF. 1).  
 FT CONFLICT 626 626 V->A (IN REF. 1).  
 SQ SEQUENCE 659 AA; 72465 MW; C2ED06A1DB8A7B81 CRC64;

Query Match 10.18; Score 275.5; DB 1; Length 659;  
 Best Local Similarity 21.68; Pred. No. 7.3e-12;  
 Matches 140; Conservative 107; Mismatches 255; Indels 147; Gaps 26;

QY 1 MRKISVIGITILLSVVLQCKSSHSSTSRGELAIN--IRDEPSRLDPQVRLLSL 58  
 Db 1 MKNRVFATAGLVLLAAGVLAAC--SSKSSDSSAPKAYGYVTADPEFLDYLSSKNST 59  
 QY 59 SLVKHYEGLVQENLNSIPALAEYSLSSDGLTYTFKLKSA---FWSNGD---PLTA 112  
 Db 60 VVTSNGIDGLTNDY--GNLAPVAEDWEVSKDGLTYTKIRKGVKVFSDGEYAEVTA 118  
 QY 113 EDFIESKQVATQEVSGYIFALNPKIKVRIKIQEG--HLSIDHFGVHPNRESTLWVLESP 171  
 Db 119 KDFVNLGHADKSEAY--LAENSVKGLADYLSGTSTDFSTGVGVKAVDDYDTLYQLNQP 177  
 QY 172 TSFLKLLALPVFPVHKSSORTLQSKSLP-----IASGAFYPRNIKOKOWIKLSKNP 223  
 Db 178 EPFWSKLTYSIFWPLNEEFETSKGSDFAKPTDPTSLYNGPFFLLKGLPAKSSVEFVKNE 237  
 QY 224 HYVNSQVETITIT-----HFIPDNTAAKLFNOCKLNWQGPWGERIPQE 270  
 Db 238 QVWKNVHLOTINLAYVYDGSQDESLEFNFTSGAYSVARLYPTSS--NY-----SKVAEE 290  
 QY 271 TLSNLSQKGLHSDVAGTSWLTFFNINKPFLN-----NMKLEALAS 312  
 Db 291 YKDNIT-----YITGSGSGLAGLVNIDROSYNVTSKTTDSEKVAATKALINKDFRQALNF 345  
 QY 313 ALDKEA-----LVSTIFL-----GRAKTADHLLPTNIHSYPEHQ-----347  
 Db 346 ALDRSAVSAQINGKDAALAVRNLFVKPDPVSAGETFGDLVAQAQIPAYGDEKGVNLA 405  
 QY 348 -QEMAQROCAVAKLFEALEELQITADLEHLNLIFFPVSSASSLL--VOLIREQWKESL 404  
 Db 406 QDGLFNADKAKAEFAKAKALEADGVQFP--IHLDPVPDQASKNYISRTQSFQSVETVL 464  
 QY 405 GFAPIV-----GKEF---ALLQADLSNGNFSIATG--GWFAFDADPMAFLTI-----447  
 Db 465 GVENVVDIQQMTSDEFNLITYYANASSEDWDVSGVSGWGDYQDPSTYDLILKTTSS 524  
 QY 448 -----FAPVSGVPYATNKHDFLEILQNTQEQ--QDHQKRSLEYLSQASLYLETFHIEP 499  
 Db 525 TKTYLGFDPNPSVYVQGLKEDYKLVDEAARETSDLVNRYEKYAAQAQWLTSSIFIP 584  
 QY 500 IYHDAFOFAMNKLNLGVSPP-----TCV-----VDFRYAK 530  
 Db 585 -----AMASSGAAPVLSRIVPFTGASAQTGSKGSDVYKYLK 621

RESULT 15  
 ALIA\_STRPN  
 ID ALIA\_STRPN STANDARD; PRT; 660 AA.  
 AC P35592; O54782; O54620; O52228;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE oligopeptide-binding protein alia precursor (Exported protein 1).  
 GN ALIA OR EXP1 OR PLPA OR SP0366.

OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_taxid-1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R800;  
 RX MEDLINE=94328326; PubMed=8051706;  
 RA Alloing G., de Philip P., Claverys J.-P.;  
 RT "Three highly homologous membrane-bound lipoproteins participate in  
 oligopeptide transport by the Ami system of the Gram-positive  
 Streptococcus pneumoniae.";  
 RL J. Mol. Biol. 241:44-58(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCIC 11906, SP-496, SP-VA92, AND SP-VA96;  
 RX MEDLINE=98125733; PubMed=9466257;  
 RA Coffey T.J., Enright M.C., Daniels M., Morona J.K., Morona R.,  
 Rykiewicz W., Paton J.C., Spratt B.G.;  
 RT "Recombinational exchanges at the capsular polysaccharide  
 biosynthetic locus lead to frequent serotype changes among natural  
 isolates of Streptococcus pneumoniae.";  
 RL Mol. Microbiol. 27:73-83(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TIGR4;  
 RX MEDLINE=21357209; PubMed=11463916;  
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
 Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,  
 Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
 Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,  
 McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
 Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
 Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
 RT "Complete genome sequence of a virulent isolate of Streptococcus  
 pneumoniae.";  
 RL Science 293:498-506(2001).  
 RN [4]  
 RP SEQUENCE OF 18-660 FROM N.A.  
 RC STRAIN=R6X;  
 RX MEDLINE=95020610; PubMed=7523829;  
 RA Pearce B., Naughton A.M., Masure H.R.;  
 RT "Peptide permeases modulate transformation in Streptococcus  
 pneumoniae.";  
 RL Mol. Microbiol. 12:881-892(1994).  
 RN [5]  
 RP SEQUENCE OF 347-509 FROM N.A.  
 RC STRAIN=R6X;  
 RX MEDLINE=95020625; PubMed=7934910;  
 RA Pearce B.J., Yin Y.B., Masure H.R.;  
 RT "Genetic identification of exported proteins in Streptococcus  
 pneumoniae.";  
 RL Mol. Microbiol. 9:1037-1050(1993).  
 CC -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM  
 FOR OLIGOPEPTIDES; PROBABLY AN OLIGOPEPTIDE BINDING PROTEIN.  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor  
 (Probable).  
 CC -!- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING  
 PROTEIN FAMILY 5.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z35135; CA84507.1; -;  
 DR EMBL; AE007348; AAK74534.1; ALT\_INIT.  
 DR EMBL; AF030359; AAC38676.1; -;  
 DR EMBL; AF030360; AAC38681.1; -;



DR EMBL; AF030361; AAC38686.1; -  
 DR EMBL; AF030364; AAC38703.1; -  
 DR EMBL; L20556; AAA26952.1; -  
 DR TIGR; SP0366; -  
 DR InterPro; IPR000914; SBP\_bac\_5.  
 DR Pfam; PF00496; SBP\_bac\_5; 1.  
 DR PROSITE; PS00013; PROKAR\_LIPOPROTEIN; 1.  
 DR PROSITE; PS01040; SBP\_BACTERIAL\_5; 1.  
 KW Peptide transport; Transport; Membrane; Lipoprotein; Signal;  
 KW Complete proteome.  
 FT SIGNAL 1 22 PROBABLE.  
 FT CHAIN 23 660 OLIGOPEPTIDE-BINDING PROTEIN ALIA.  
 FT LIPID 23 23 N-ACYL DIGLYCERIDE (PROBABLE).  
 FT VARIANT 7 7 F -> L (IN STRAIN R800).  
 FT VARIANT 27 27 G -> D (IN STRAINS NCTC 11906, SP-VA92,  
 FT SP-496 AND SP-VA96).  
 FT VARIANT 30 30 T -> A (IN STRAINS R800 AND R6X).  
 FT VARIANT 166 166 E -> D (IN STRAINS NCTC 11906, SP-VA92,  
 FT SP-496 AND SP-VA96).  
 FT VARIANT 246 246 V -> I (IN STRAINS R800 AND R6X).  
 FT VARIANT 368 368 L -> I (IN STRAINS NCTC 11906, SP-VA92,  
 FT SP-496 AND SP-VA96).  
 FT VARIANT 432 432 Q -> T (IN STRAINS R800, R6X, NCTC 11906,  
 FT SP-VA92, SP-496 AND SP-VA96).  
 FT VARIANT 559 559 T -> A (IN STRAINS NCTC 11906, SP-VA92,  
 FT SP-496 AND SP-VA96).  
 FT VARIANT 612 612 T -> A (IN STRAIN SP-496).  
 FT CONFLICT 18 19 TT -> GV (IN REF. 4).  
 FT CONFLICT 137 137 L -> P (IN REF. 4).  
 FT CONFLICT 420 420 A -> R (IN REF. 1).  
 SQ SEQUENCE 660 AA; 73079 MW; 70255692EC05B8 CRC64;

Query Match 9.9%; Score 272.5; DB 1; Length 660;  
 Best Local Similarity 21.4%; Pred. No. 1.2e-11;  
 Matches 128; Conservative 105; Mismatches 229; Indels 137; Gaps 24;

QY 12 ILISLSVLQCKESHSSSTSRGELAINIRDEPRSLDPQVRLLSEISLVKHYEGLYOE 71  
 DB 12 VILLAAATLACSGSGSTGKGTFTSYIYETDPDLNLTAKAANTANITSNVVDGLL-E 70  
 QY 72 NMLSGNIIPALAEYLSLSDGLTYFKL-KSAFW--SNGD---PLTAEDFIESWKQVATQ 125  
 DB 71 NDYGNFVPSMAEDWSKDGLTYYTIRKDAKWTSEGEYAAVRAQDFVTGLKYAADK 130  
 QY 126 EYSGIYAFALNPKNKVKIQEHL-SIDHFGVHSPNESTLVVLTESPTSHFLKILLALPVF 184  
 DB 131 KSDALY-LVQESINGLDAYVKGIEKDFQVGIKALDEQTVQYTLNKPESFWSKTTMGVL 189  
 QY 185 FVHKSQRTLOSK-----SLPIASGAFYKPNIKQKQWIKLSKNPHYNOQSVETK 234  
 DB 190 APV--NEEFLNSKDDDFAKATDPSSLLYNGPYLLKSIVTKSSVEFAKPNYWDKNVHD 247  
 QY 235 TITHTFT--PDANTAALFNQGLKNWQPPWGERI-----AARLYPTSASFAELEKSKMKNIVYTQDS 271  
 DB 248 KYKLSFWGDGDTSKPAENFKDGLT-----AARLYPTSASFAELEKSKMKNIVYTQDS 301  
 QY 272 L-----SNLQSKGHLHSDVAGTSHLFTFNKFPNNMKLREALASALDKEALYS----- 321  
 DB 302 IYLVGINIDRQSYKT-----SKTDEQKASTKALLNKDFQAIAPGFDRTAYASQLNGQ 358  
 QY 322 -----TIFLGRAKTADHLPTNIHSYPEHOK-----QEMAQROAYAKKL 360  
 DB 359 TGASKILRNLFVPTFVQADGKFCMDVKEKLVYGDENKVDNLDASQDGLYNPEKAKAE 418  
 QY 361 FKEALELQITAKDLH-LNLIFFVSSASLL--VQLIREQWKSGLG----- 405  
 DB 419 FAKAKSALQ--AEGVQFPIHLDMPVDQATTQVQRVQSMKQSLATLGADNVIIDIOQLQ 476  
 QY 406 -----FAIPIVKGFAALLQDLSSGNFSLATGCGFADFPMAFL----- 445  
 DB 477 KDEVNNITYFAENAGEDW-----DLSD-----NVGNGPDPADPSTVLDIIRKPSVGEST 525

QY 446 -TIFAYPSG---VPPYAINHKDFLEILQNT-BQEOOHOKRSELYSQASLYLETFHIEP 499  
 DB 526 KTVLGFDSGEDNVAKKVGLDYDEKLYTEAGDETTDVAKRYDKYAAAAQAWLTDSALIIP 584

Search completed: July 26, 2002, 04:42:32  
 Job time: 260 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 26, 2002, 04:36:22 ; Search time 87.81 seconds  
(without alignments)  
1048.096 Million cell updates/sec

Title: US-09-824-567-2

Perfect score: 2739

Sequence: 1 MKRISVIGITILLSVVL.....LSNLGVSPGWDPRYAKEN 532

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Lasting first 45 summaries

Database :

SPTREMBL\_19:\*

- 1: sp.archaea:\*
- 2: sp.bacteria:\*
- 3: sp.fungi:\*
- 4: sp.human:\*
- 5: sp.invertebrate:\*
- 6: sp.mammal:\*
- 7: sp.mhc:\*
- 8: sp.organelle:\*
- 9: sp.phage:\*
- 10: sp.plant:\*
- 11: sp.rodent:\*
- 12: sp.virus:\*
- 13: sp.vertebrate:\*
- 14: sp.unclassified:\*
- 15: sp.virus:\*
- 16: sp.bacteriophage:\*
- 17: sp.archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2739	100.0	532	16 Q928Y9	Q928Y9 chlamydia p
2	1364	49.8	518	16 Q84201	Q84201 chlamydia t
3	1345	49.1	520	16 Q9PKJ4	Q9PKJ4 chlamydia m
4	1160.5	42.4	527	16 Q928Y8	Q928Y8 chlamydia p
5	755.5	27.6	526	16 Q928Y6	Q928Y6 chlamydia p
6	755.5	27.6	528	16 Q9K244	Q9K244 chlamydia p
7	748	27.3	527	16 Q9PKL9	Q9PKL9 chlamydia m
8	685.5	25.0	529	16 Q84178	Q84178 chlamydia t
9	536.5	19.6	550	16 Q97046	Q97046 clostridium
10	509.5	18.6	528	2 Q9K5F7	Q9K5F7 borrelia ga
11	506	18.5	528	2 Q9K5J9	Q9K5J9 borrelia af
12	505.5	18.5	528	2 Q9K5J9	Q9K5J9 borrelia bu
13	505.5	18.5	528	16 Q54584	Q54584 borrelia bu
14	482.5	17.6	435	16 Q928Y7	Q928Y7 chlamydia p
15	479.5	17.5	532	16 Q9AKR0	Q9AKR0 rhizobium m
16	478	17.5	559	16 Q929H6	Q929H6 listeria in

17	476	17.4	550	2 Q93QH8	Q93QH8 lactococcus
18	474.5	17.3	558	2 Q9LAT7	Q9LAT7 listeria mo
19	472	17.2	550	16 Q9CIL3	Q9CIL3 lactococcus
20	471.5	17.2	530	16 Q31315	Q31315 borrelia bu
21	467.5	17.1	514	2 Q53480	Q53480 borrelia co
22	467	17.1	545	16 Q9CIL2	Q9CIL2 lactococcus
23	446.5	16.3	553	16 Q927S4	Q927S4 listeria in
24	446	16.3	551	2 Q9F5U6	Q9F5U6 bacillus th
25	446	16.3	565	16 Q97D48	Q97D48 clostridium
26	441	16.1	529	2 Q9F1J4	Q9F1J4 enterococcu
27	440.5	16.1	519	2 Q9RMW8	Q9RMW8 bacillus an
28	431	15.7	543	16 Q9KTI4	Q9KTI4 vibrio chol
29	429.5	15.7	523	2 Q31313	Q31313 borrelia bu
30	428.5	15.6	529	2 Q31303	Q31303 borrelia bu
31	428.5	15.6	529	16 Q50927	Q50927 borrelia bu
32	428.5	15.6	530	16 Q985N6	Q985N6 rhizobium l
33	427.5	15.6	523	2 Q52615	Q52615 borrelia bu
34	425.5	15.5	529	16 Q51307	Q51307 borrelia bu
35	421	15.4	549	16 Q92FA6	Q92FA6 listeria in
36	419	15.3	426	16 Q84141	Q84141 chlamydia t
37	415.5	15.2	541	2 Q52616	Q52616 borrelia bu
38	414.5	15.1	541	16 Q51308	Q51308 borrelia bu
39	412	15.0	428	16 Q9PKP9	Q9PKP9 chlamydia m
40	412	15.0	549	16 Q9C127	Q9C127 lactococcus
41	411.5	15.0	523	2 Q31304	Q31304 borrelia bu
42	409.5	15.0	541	2 Q31306	Q31306 borrelia bu
43	400.5	14.6	547	16 Q83594	Q83594 treponema p
44	390.5	14.3	545	2 Q51643	Q51643 enterococcu
45	379	13.8	522	2 Q06515	Q06515 streptococcu

#### ALIGNMENTS

RESULT 1

Q928Y9 PRELIMINARY; PRT; 532 AA.

AC Q928Y9; 1000 (TREMBLrel. 10, Created)

DT Q928Y9; 1000 (TREMBLrel. 10, Last sequence update)

DT Q928Y9; 1000 (TREMBLrel. 19, Last annotation update)

DE OLIGOPEPTIDE BINDING PROTEIN.

GN OPPA1 OR CPN0195 OR CP0572.

OS Chlamydia pneumoniae (Chlamydia pneumoniae).

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI\_TaxID=83558;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CWL029.

RX MEDLINE=99206606; PubMed=10192388;

RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,

RT Olinger L., Grimwood J., Davis R.W., Stephens R.S.;

RL "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";

RL Nat. Genet. 21:385-389(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=AR39;

RX MEDLINE=20150255; PubMed=10684935;

RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,

RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,

RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,

RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,

RA Eisen J., Fraser C.M.;

RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia

RT pneumoniae AR39.";

RL Nucleic Acids Res. 28:1397-1406(2000).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=J138;

RX MEDLINE=20330349; PubMed=10871362;

RA Shirai M., Hiraoka H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,

RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;

RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138

```

RT from Japan and CWL029 from USA.;
RL Nucleic Acids Res. 28:2311-2314(2000).
DR EMBL; AE001606; AAD18348.1; -
DR EMBL; AE002216; AAP38391.1; -
DR EMBL; AP002545; BAA98405.1; -
DR TIGR; CP0572; -
DR InterPro; IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5; 1.
KW Complete proteome.
SQ SEQUENCE 532 AA; 59744 MW; 1CB473D9D46A1579 CRC64;

Query Match      100.0%; Score 2739; DB 16; Length 532;
Best Local Similarity 100.0%; Pred. No. 1.5e-196;
Matches 532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKISVGICITILLSSVVLQCKESSHSSTSGELAINIRDEPRSLDPRQVRLLEISL 60
Db 1 MKKISVGICITILLSSVVLQCKESSHSSTSGELAINIRDEPRSLDPRQVRLLEISL 60
QY 61 VKHIYEGVLQENNSGNIPEALAEYSLSSDGLTYTFKLKSAFWSNGDPLTAEDFIESWK 120
Db 61 VKHIYEGVLQENNSGNIPEALAEYSLSSDGLTYTFKLKSAFWSNGDPLTAEDFIESWK 120
QY 121 QVATQEVSGIYAPALNPKNVKIKQEGHLSIDHFGVHSPNESTLVYLTLESPTSHFLKLLA 180
Db 121 QVATQEVSGIYAPALNPKNVKIKQEGHLSIDHFGVHSPNESTLVYLTLESPTSHFLKLLA 180
QY 181 LPVFFPHKRSQRTLSKSLPIASGAFTPKNIKOKWIKLSKNPHYNSQVETKTIITHF 240
Db 181 LPVFFPHKRSQRTLSKSLPIASGAFTPKNIKOKWIKLSKNPHYNSQVETKTIITHF 240
QY 241 IPDNTAAKLFNQGKLNQWGPWGERIPOETLSNLSQKGLHSFVAGTSMLTENINKFP 300
Db 241 IPDNTAAKLFNQGKLNQWGPWGERIPOETLSNLSQKGLHSFVAGTSMLTENINKFP 300
QY 301 LNNMKLEALASALDKEALVSTIFLGRAKTADHLLPTNTHSYPEHQKQMAQRAYAKKL 360
Db 301 LNNMKLEALASALDKEALVSTIFLGRAKTADHLLPTNTHSYPEHQKQMAQRAYAKKL 360
QY 361 FKALELEQITAKDLEHLNLIFFPVSSASSLLVQIREQWKSGLFAIPVIGKEFALLQA 420
Db 361 FKALELEQITAKDLEHLNLIFFPVSSASSLLVQIREQWKSGLFAIPVIGKEFALLQA 420
QY 421 DLSSGNFSLATGWADFADPAFLTIIPAYPSGVPYPAIYNNKDFLEILQNEQEQDQHR 480
Db 421 DLSSGNFSLATGWADFADPAFLTIIPAYPSGVPYPAIYNNKDFLEILQNEQEQDQHR 480
QY 481 SELVSOASLYLETHIEPIYHDAFOFAMNKKLSNLGVSPTGVVDYFVAKEN 532
Db 481 SELVSOASLYLETHIEPIYHDAFOFAMNKKLSNLGVSPTGVVDYFVAKEN 532

RESULT 2
ID 084201 PRELIMINARY; PRT; 518 AA.
AC 084201;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DE OLIGOPEPTIDE BINDING PROTEIN.
GN OPPA_3 OR CT198.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UV-3/CX;
RX MEDLINE=99008080; PubMed=9784136;
RA Stephens R.S., Kalman S., Lamell C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:

Chlamydia trachomatis.";
RL Science 282:754-759(1998).
DR EMBL; AE001293; AAC67790.1; -
DR InterPro; IPR000914; SBP_bac_5.
DR Pfam; PF00496; SBP_bac_5; 1.
KW Complete proteome.
SQ SEQUENCE 518 AA; 58858 MW; 8B3AE840831BBEF1 CRC64;

Query Match      49.8%; Score 1364; DB 16; Length 518;
Best Local Similarity 50.1%; Pred. No. 9.8e-94;
Matches 269; Conservative 97; Mismatches 147; Indels 24; Gaps 8;

QY 1 MKKISVGICITILLSSVVLQCKESSHSSTSGELAINIRDEPRSLDPRQVRLLEISL 56
Db 1 MKKISVGICITILLSSVVLQCKESSHSSTSGELAINIRDEPRSLDPRQVRLLEISL 56
QY 57 ELSLVKHIYEGVLQENNSGNIPEALAEYSLSSDGLTYTFKLKSAFWSNGDPLTAEDFI 116
Db 57 DINLHLHLYEGLVQETP-SGEVFPALAESEFFLSEDKKTYTFNLKKAFFWSNGDLTAHDFV 115
QY 117 ESWKQVATQEVSGIYAPALNPKNVKIKQEGHLSIDHFGVHSPNESTLVYLTLESPTSHFL 176
Db 116 RSWNDVLQNRVASYISFAFLPI-DVWK-----DSGFFAKDDHLLVILNLTPTPHFL 165
QY 177 KLLALPVFFPHV-KSORTLSQSKSLPIASGAFTPKNIKOKWIKLSKNPHYNSQVETKT 235
Db 166 KLLTLPVFPVHSQHQIRKEEKSLSPISTGAFFLKEKDRWLKLEKSPYNNKQVAVQE 225
QY 236 ITIHFIPDNTAAKLFNQGKLNQWGPWGERIPOETLSNLSQKGLHSFVAGTSMLTEN 295
Db 226 ICIIHIIQQDTAGALFNQGLDWGLPMGHSIPQETLATTNKRRAPRFSDISGTSWLTEN 285
QY 296 INKFLNNMKLEALASALDKEALVSTIFLGRAKTADHLLPTNTHSYPEHQKQMAQRAYAKKL 355
Db 286 TAKPFSSHLSKRLQALSILNKEALSAFV---KPAKHLPAHLHTYPEQSPYKQQAET 342
QY 356 YAKKLFKALELEQITAKDLEHLNLIFFPVSSASSLLVQIREQWKSGLFAIPVIGKEF 415
Db 343 LAKSLLEALTELNMTIEDLEKYPLTFSTATSNWSQIAQMLRDQWRRSLGTFPPICGKEY 402
QY 416 ALQALDSSGNFSLATGWADFADPAFLTIIPAYPSGVPYPAIYNNKDFLEILQNEQEQD 475
Db 403 ALLQNDLIGNTFMISGWFADPSDPLAFSLIFS-SKGVKPYALQDPQFQDLILSLETEK 461
QY 476 DHKRSYLSQASLYLETHIEPIYHDAFOFAMNKKLSNLGVSPTGVVDYFVAKEN 532
Db 462 NPKRSALISEASLVIERQNVIEPLYHDVPHYTNKLSFVFLHPSGLVDVRYAKNS 518

RESULT 3
QYKJ4
ID Q9PKJ4 PRELIMINARY; PRT; 520 AA.
AC Q9PKJ4;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE PEPTIDE ABC TRANSPORTER, PERIPLASMIC PEPTIDE-BINDING PROTEIN,
GN PUTATIVE.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / NIGG;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia

```



366 cctggacggactcacttatacttta 392

306 ttcaagaaaaataatctttcaggaaatatagagcctgctcttcagagaagactactctctt 365



```

Query Match      2.0%; Score 35.8; DB 1; Length 7218;
Best Local Similarity 5.4%; Pred. NO. 1;
Matches 22; Conservative 205; Mismatches 182; Indels 0; Gaps 0;

QY 338 cctgccttgcagaagactctcttctctcgacggactcaattatacttttaactg 397
      :: :::: : ::::: ::::: : ::::: : ::::: : ::::: : ::::: :
Db 1072 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1131

QY 398 aaatcagcttttggagtaatggcgacccttaacgctgaagactttatagaactctg 457
      :: ::::: : ::::: : ::::: : ::::: : ::::: : ::::: :
Db 1132 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1191

QY 458 aaacaagtactactcaagaagctcaggaaactatgcttgccttgaaatccaaataa 517
      :: ::::: : ::::: : ::::: : ::::: : ::::: : ::::: :
Db 1192 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1251

QY 518 aatgtacgaagatccaagagggacactctccatagaccatttggagtcactctct 577
      :: ::::: : ::::: : ::::: : ::::: : ::::: : ::::: :
Db 1252 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1311

QY 578 aatgaattcacactgtgttaacctggaaatcccaacctgcattctctaaactttta 637
      :: ::::: : ::::: : ::::: : ::::: : ::::: : ::::: :
Db 1312 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1371

QY 638 gctcttcacgtcttttcccgcttcataaactccaaagaaacctgcaatccaaactcta 697
      ::::: : ::::: : ::::: : ::::: : ::::: : ::::: :
Db 1372 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1431

QY 698 cctatagcagcgagagcttctctactctaaataatccaacaaaacaaat 746
      ::::: : ::::: : ::::: : ::::: : ::::: : ::::: :
Db 1432 YYGTACCAAAATTCCTATCTCTTTAACTACTTGCATAGATAGCTAAT 1480

RESULT 7
US-08-687-080-57/C

```

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: INTRON 1 OF RAD50 GENOMIC SEQUENCE
US-08-687-080-57

Query Match 1.9%; Score 35; DB 2; Length 1829;
Best Local Similarity 56.5%; Pred.No.0.85;
Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1477 caaccataagagctctctagaattctcaaaacatagacaagagcaaggtaccacaaa 1536
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 803 CAATCCTGAGTACATACAGACATTCAACAATAACATTTTACAAAAACAAAGTAAC 744

QY 1537 acgctcggaattagtcgcaagctctcttacctagagaccttcattatt 1591
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 743 AAGTTTGTCATTCATTCTCAACCTTCTCTTCTTCTACACTCTTCACAAATCCTT 689

RESULT 8
US-09-381-862-1
; Sequence 1, Application US/09381862
; Patent No. 6245906
; GENERAL INFORMATION:
; APPLICANT: Ueyama, Hiroshi
; APPLICANT: Abe, Kanako
; APPLICANT: Keshi, Hiroyuki
; APPLICANT: Matsuhsa, Akio
; TITLE OF INVENTION: PROBES FOR THE DIAGNOSIS OF INFECTIONS
; TITLE OF INVENTION: CAUSED BY STREPTOCOCCUS PYOGENES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun

```



STREET: 233 South Wacker Drive/6300 Sears Tower  
 CITY: Chicago  
 STATE: Illinois  
 COUNTRY: United States of America  
 ZIP: 60606

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/381,862  
 FILING DATE:

## CLASSIFICATION: 536

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 1997-71077  
 FILING DATE: 25-MAR-1997

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP98/01288  
 FILING DATE: 23-MAR-1998

## ATTORNEY/AGENT INFORMATION:

NAME: Cawley, Jr., Thomas A.  
 REGISTRATION NUMBER: 40,944  
 REFERENCE/DOCKET NUMBER: 19036/36274  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (312) 474-6300  
 TELEFAX: (312) 474-0448

## INFORMATION FOR SEQ ID NO: 1:

## SEQUENCE CHARACTERISTICS:

LENGTH: 3549 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

## ORIGINAL SOURCE:

ORGANISM: Streptococcus pyogenes  
 STRAIN: Clinical Isolate SP-6-28  
 US-09-381-862-1

## Query Match

Best Local Similarity 1.9%; Score 35; DB 4; Length 3549;

Matches 56; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 306 ttcaagaaataattcttcaggaaatagacctgtcttcgagaagactactctcttt 365

Db 2486 TTTTGGAAACGACGATATGGTATTTAGTACCATCTTGCTAAAGATTGGAAGGTTT 2545

Qy 366 cctcgagcggactcaattatacttttaact 396

Db 2546 CAAAAGAGGCTGACTATATACCTATACCT 2576

## RESULT 9

US-09-824-567-1

Sequence 63, Application US/08545860D

Patent No. 6040140

## GENERAL INFORMATION:

APPLICANT: Croce, Carlo

APPLICANT: Cavaani, Eli

TITLE OF INVENTION: Diagnostics, Therapeutics and Methods

TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias

TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region

NUMBER OF SEQUENCES: 94

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &

ADDRESSEE: No. 6040140ris

STREET: One Liberty Place, 46th floor

CITY: Philadelphia

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19103

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/545,860D  
 FILING DATE: 07-MAR-1996  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US94/04496  
 FILING DATE: 22-APR-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US92/10930  
 FILING DATE: 09-DEC-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/327,392  
 FILING DATE: 19-OCT-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/320,559  
 FILING DATE: 11-OCT-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/062,443  
 FILING DATE: 14-MAY-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/971,094  
 FILING DATE: 30-OCT-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/888,839  
 FILING DATE: 27-MAY-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/805,093  
 FILING DATE: 11-DEC-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Deluca Esq., Mark  
 REGISTRATION NUMBER: 33,229  
 REFERENCE/DOCKET NUMBER: TLU-1262  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 568-3100  
 TELEFAX: (215) 568-3439  
 INFORMATION FOR SEQ ID NO: 63:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 8342 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 2..265  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 595..666  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 2353..2484  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 3032..3145  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 6788..6934  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 7967..8062  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 8304..8342  
 US-08-545-860D-63

## Query Match

1.9%; Score 34.8; DB 3; Length 8342;

```

; NAME/KEY: CDS
; LOCATION: 3032..3145
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 6788..6934
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 7967..8062
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 8304..8342
;
PCT-US94-04496-63

Query Match      1.9%; Score 34.8; DB 5; Length 8342;
Best Local Similarity 49.5%; Pred. No. 2.3;
Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 658 cgttcataaattcacaagaacccgtgcaatccaatctcactatagcaagcgagcttt 717
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3302 CCTTCTTAATCTCTAATACTTCAAAACATTTCTTAGCTGGTTTAAACAGGGAATA 3243

QY 718 ctatctctaaaaatatcaaaacaaaacaatggataaaactctcaaaaaacccctcactacta 777
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3242 GTACCTGCTAAATAACCAACAAACAGACTATAGTAATCTCATCCCCAGATCAATCATTTA 3183

QY 778 taatcaagtcaggtggaaactaaacagattacgattcactcattcccgatgcaaacac 837
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3182 TGTACTTCTCGGTGACAAAGAGCAGGATGCCTTACCTCTACATGCCCATCTACTGGCAC 3123

QY 838 ag 839
      ||
DB 3122 AG 3121

RESULT 11
US-09-450-852-3
; Sequence 3, Application US/09450852
; Patent No. 6309860
; GENERAL INFORMATION:
; APPLICANT: Chaganti, Raju S.K.
; APPLICANT: Dyomin, Vadim
; APPLICANT: Dalla-Favera, Riccardo
; TITLE OF INVENTION: CLONING AND USES OF BCL-8
; FILE REFERENCE: 53828-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/450,852
; CURRENT FILING DATE: 1999-11-29
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2340
; TYPE: DNA
; ORGANISM: human
US-09-450-852-3

Query Match      1.9%; Score 34.4; DB 4; Length 2340;
Best Local Similarity 60.9%; Pred. No. 1.5;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1668 ttgtggacttcggttatgctaaggaaaattagcaactcttttaactcgcgaacttgtaa 1727
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1970 ttattgactacactattctgaggctactggcaccctcttctgatttcatactattgtaa 2029

QY 1728 agaactgaattcttataactaaactgggtgcctt 1759
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2030 tgacattagatatttcttatattgtctgcatt 2061

RESULT 12
US-08-617-860B-3/c
; Sequence 3, Application US/08617860B
; Patent No. 6133506

```

```

; NAME/KEY: CDS
; LOCATION: join(2611..2908, 3001..3341)
US-08-617-860B-3

```

[illegible]

RESULT 14  
US-08-378-580-5/c  
; Sequence 5, Application US/08378568  
; Patent No. 5608148  
; GENERAL INFORMATION:  
; APPLICANT: John, Maliyakal E.  
; TITLE OF INVENTION: TRANSGENIC COTTON PLANTS  
; TITLE OF INVENTION: PRODUCING HETEROLOGOUS PEROXIDASE  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady

Search completed: July 26, 2002, 03:36:33  
Job time: 2316 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 02:57:07 ; Search time 1946.51 Seconds  
(without alignments)  
12474.132 Million cell updates/sec

```

Title: US-09-824-567-1
Perfect score: 1799
Sequence: 1 acctccccctgctaaacta.....tactgctctctctctcta 1799

```

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched:	13736207 seqs, 6748477542 residues
Total number of hits satisfying chosen parameters:	27472414

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

Database : EST: \*

```

1:  est: *
2:  em_estba: *
3:  em_esthum: *
4:  em_estmu: *
5:  em_estov: *
6:  em_estpl: *
7:  em_estro: *
8:  em_htc: *
9:  gb_est1: *
10: gb_est2: *
11: gb_htc: *
12: gb_gss: *
13: em_gss_hun
14: em_gss_inv
15: em_gss_plr
16: em_gss_vir

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	%			Query Match	Length	DB	ID	Description
	Score	Match						
1	53.4	3.0		1101	12	CNS00L00	AL068607	Drosophil
C 2	51.2	2.8		1101	12	CNS00L72	AL078714	Drosophil
3	50.2	2.8		1101	12	CNS017HL	AL108051	Drosophil
C 4	48.5	2.7		1031	12	CNS00CF2	AL059199	Drosophil
5	48.6	2.7		1101	12	CNS0039G	AL063921	Drosophil
6	47.4	2.6		939	12	CNS00CNG	AL059400	Drosophil
C 7	46	2.6		1101	12	CNS0100X	AL098379	Drosophil
8	45.6	2.5		1101	12	CNS0163Y	AL106840	Drosophil
9	45.2	2.5		859	12	CNS0047Y	AL055406	Drosophil
C 10	45.2	2.5		1201	12	CNS0168G	AL106426	Drosophil
C 11	44.6	2.5		1101	12	CNS00051	AL065414	Drosophil
C 12	44.2	2.5		1101	12	CNS00FO0	AL070854	Drosophil
13	43.8	2.4		555	12	CNS040CR	AL298444	Tetradon
14	43.8	2.4		570	10	BI209849	EST527889	
C 15	43.8	2.4		922	12	CNS0073W	AL066784	Drosophil
C 16	43.8	2.4		1101	12	CNS016K3	AL106845	Drosophil
C 17	43.4	2.4		1101	12	CNS00EP6	AL063475	Drosophil

## SUMMARIES

RESULT	1	
CNS00LO0		
LOCUS		
DEFINITION		
BACR32D23 of RPCI-98 library from <i>Drosophila melanogaster</i> (fruit fly), genomic survey sequence.		
ACCESSION		
VERSION		
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		

REFERENCE AUTHORS TITLE JOURNAL	COMMENT
MUSCOPHOPHINA; EPYNAUTOIDEA; DROSOPHILIDAE; DROSOPHILA. 1 (bases 1 to 1101) Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - web : www.genoscope.cns.fr)	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pletier de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPOJ-98 and was constructed by partial, EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila">http://bacpac.med.buffalo.edu/drosophila</a> bac.htm.

FEATURES	SOURCE
Location/Qualifiers	1. .1101
round at <a href="http://bacpac.med.builaro.edu/drosophila">http://bacpac.med.builaro.edu/drosophila</a>	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone_lib="RPOI-98"

## ALIGNMENTS

RESULT	1
CNS00LOO	
LOCUS	CNS00LOO linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR32D23 of RPCT-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL068607
VERSION	AL068607.1 GI:4958689
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 1101).

REFERENCE AUTHORS TITLE JOURNAL	COMMENT
MUSCOPHOPHINAE, EPYDROPHILINAE; DROSOPHILINAE; DROSOPHILA. 1 (bases 1 to 1101) Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - web : www.genoscope.cns.fr)	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pictet de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPQ1-98 and was constructed by partial, EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at <a href="http://bacpac.med.buffalo.edu/drosophila">http://bacpac.med.buffalo.edu/drosophila</a> bac.hum.

FEATURES	SOURCE
Location/Qualifiers	1. .1101
round at <a href="http://bacpac.med.builaro.edu/drosophila">http://bacpac.med.builaro.edu/drosophila</a>	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/clone_lib="RPOI-98"







[illegible]

	QY	B14	tacttcaattccgcatgcgaacagcagcgacaatctatttaaccggaggaaacctcaattg	877			
	Dd	B01	AHHNHUYCTMHNHMTBHHCNCHCNYTMHNNHYTUTAAAMMAAAAAAAMAATAA	942			
	QY	B74	gsaaggaccttccttggtggagaacsuccatctccaagaacccttatccaatttaacagtcta	933			
	Dd	B41	AAAAG-----NAAGAATGATTGTATTATTATTTTTCCTTTTTTAATWTYTWTHTCMNS	696			
	QY	B34	ggggcacttacactcttttgtatgccgaggaacctcaggtcacctcctcaatatcaaataa	993			
	Dd	B65	HCHCCACSMNVHNCYCUNYMNMHMHWKCNWAKASADANMMVMAAMVMWCSCMCWC	636			
	QY	B94	attccccctcaaatgatgaagcttagadaagccttagcatcgctcagcttagataaggagc	1052			
	Dd	B35	WCYUXTSAMTHNNMHMNHNNTTNHANHAANTGAACAASCWAATHATMCTHNTMNTT	576			
	QY	B054	tttgtctoaactatattcttagccgtgcgaacaaactgcgcgcatctcctcactcataa	1112			
	Dd	B75	YHTUTCAYTGTYWHCWCAHANYKCYNMMMACATYAMAMAHANAATMYTWMTWMHNAWSC	516			
	QY	B114	tattoatagatctcccgaactcaaaaaagadagatgcgcaacgccaagcttagcgttaa	1172			
	Dd	B15	TNHTWBHAMHANHMANMWNHAWAABACHNNMAMAMAHANNMWCSAALAWAWARAWA	456			
	QY	B174	aaaactctttaagaagccttagaagaacctccaaatacac	1212			
	Dd	B55	HAWHBNTMCWANMAMHANHAYAHAWMTTCMCCHOCHC	417			
RESULT 8							
CNS016JY	LOCUS	Drosophila melanogaster genome survey sequence v7 end of BAC ACNI16D23 of DROSBC library from Drosophila melanogaster (fruit fly), genomic survey sequence.					
ACCESSION	VERSION	AL106840.1 GI:5624081 GSS.					
SOURCE	KEYWORDS	fruit fly.					
ORGANISM		Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Epiphydroidea; Drosophilidae; Drosophila.					
REFERENCE		1 (bases 1 to 1101) Genoscope.					
AUTHORS		Direct Submission					
TITLE	JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequen-					
COMMENT		BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscu.cns.fr) - Web : www.genoscu.cns.fr Determination of this BAC-end sequence was carried out as part collaboration with the European Drosophila Genome Project (EDGP) http://www.edgp.ebi.ac.uk . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaut at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Buche and Genevieve Payan. It has been constructed in the vector pBelOBAC11.  					
FEATURES							
source							
Location/Qualifiers							
1..1101 /organism="Drosophila melanogaster" /plasmid="pbelOBAC11" /db_xref="taxon:7227" /clone_lib="DrosBAC" /clone="BACN16D23" /note="end : 17"							
BASE COUNT	337 a	71 c	48 g	61 t 584 others			
ORIGIN							
Query Match	2.5%	Score 45.6;	DB 12;	Length 1101;			
Best Local Similarity	21.4%;	Pred. No. 5;	Matches 82;	Conservative 128;			
			Mismatches 173;	Indels 0;			
Gaps							

```

QY 481 ctccaggaatctctgtcttgcctgaatccaaattaaaaatgtacgaagaatccaaagg 540
Db 700 YYYAYYYYCYVAYYYYCYWYCYCAAYYYYCYYYAAMAAANBCYYAABCMTCABAAAA 759
QY 541 acacctctccatagacacatttggagtgcaactctccatgaatctacacttggttgac 600
Db 760 BCBAAAAAAMAAAMAACTMAVYBYBAWYAAAAAAYAAAAAYAAAAAYAAAAAYAAAA 819
QY 601 cctggaatccccaacactcgactttttaaacttttagctctccagctttttccocgt 660
Db 820 AAHCAAAHCTCTAAAAAAYYYYAAYCTYCAAAWYCYCAAYCYCAAYCYCC 879
QY 661 tcataaatctcaagaacctgcaatctctacactctacactatagcagcgagcttcta 720
Db 880 WWAYAAAYCAAYCYCHYHYHATYHYAAWYMAHAHAAYHYHAYHYHAYHYHAYHYHAY 939
QY 721 tctaaataatcaacaaacaaatgataaaactctcaaaacccctcactactataa 780
Db 940 YYYAHHAAAAAAYYYYVYAAWYAAVYAAVYAAVYAAVYAAVYAAVYAAVYAAVYAA 999
QY 781 tcaagtcaggtggaactaaacgattacgattcactcattccocgattgcaacacagc 840
Db 1000 AHMHYIAYYYYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 1059
QY 841 agcaaaactatttaacaggaa 863
Db 1060 AABATHYABANAAAAABCA 1082

RESULT 9
CNS004YY CNS004YY 859 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR11F03 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL0553406
VERSION AL0553406.1 GI:4932207
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 859)
Genoscope.
AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org the BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
1. 859
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR11F03"
/note="end : TET3"
BASE COUNT 302 a 32 c 15 g 124 t 386 others

QY 1434 taacgatcttctgtatccatcaggagttctctctatgcaatcaaccatgaagcctcc 1493

```

## ORIGIN

```

Query Match 2.5%; Score 45.2; DB 12; Length 859;
Best Local Similarity 20.6%; Pred. No. 6;
Matches 54; Conservative 104; Mismatches 104; Indels 0; Gaps 0;

QY 611 caaacctgcattcttaaaacttttagctctccagctttttccocgttcaaatat 670
Db 596 CWHAWCTCYCAYMTMAAACCYATMCCACYACYCCYMYCYCYCYCYCYCYCYCYCY 655
QY 671 caaagaacctgcaatccaaatctctacatctacatagcaacgagagcttctctct 730
Db 656 HAAAYHCTAAAHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHY 715
QY 731 atcaacaaacaaatgataaaactctcaaaacccctcactactataatcaagtcag 790
Db 716 YYAYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHY 775
QY 791 gtggaactaaacgattacgattcactcattcccgatgcaacacagcagcaaaata 850
Db 776 AAYAYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHYHY 835
QY 851 tttaatcagggaacactcaatt 872
Db 836 HWYHAAAAAHHAAHAYHY 857

```

## RESULT 10

```

CNS0168G/C CNS0168G 1201 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION BACN15114 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL106426
VERSION AL106426.1 GI:5621843
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1201)
Genoscope.
AUTHORS Direct Submission
TITLE Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pbeloBAC11.
FEATURES
Location/Qualifiers
source
1. 1201
/organism="Drosophila melanogaster"
/plasmid="pbeloBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN15114"
/note="end : T7"
BASE COUNT 337 a 214 c 258 g 159 t 233 others
ORIGIN

```

## Query Match

```

Best Local Similarity 28.1%; Score 45.2; DB 12; Length 1201;
Matches 87; Conservative 80; Mismatches 143; Indels 0; Gaps 0;

```

QY	545	ctctccatagaccattttggagtgcaactctcctaataaacctacacttgttgtacctcg	604
DG		: : :   : : : :   : : : :   : : : :   : : : :   : : : :   : : : :	
Db	1099	mtmnmymtmnmnmnhtntmmtnmnmntmtmttttttnttntnnmtttntmttgcwy	1040
QY	605	gaatccccaaacctgcgatttcttaaaaaacttttagctctcttcagctcttttcccgcgtcat	664
DG		: : : :   : : : :   : : : :   : : : :   : : : :   : : : :   : : : :	
Db	1039	mcsnnnnnnnammmmscmuytuymumnmnmnyhusct-yyuuuyyuumtanaahwamsmch	981
QY	665	aaactccaagaacctcgaaatccaaaactctcaactatagcaagcgagctttctatct	724
DG		: : : :   : : : :   : : : :   : : : :   : : : :   : : : :   : : : :	
Db	980	atrtntnaunmncsmnmnmchnstutamtumcmkscwmayuanynhasumnmnmntmn	921
QY	725	aaaaatacaaaaacaataaggataaaaactctcaaaaaaacctccactactaatcaa	784
DG		:    : : : :   : : : :   : : : :   : : : :   : : : :   : : : :   : : : :	
Db	920	nwtatsttanhammamnmntanhnmmmmnmhtscuytuymumnyuuuyyuuwtw	861
QY	785	agtcaaggtggaactaaaacgatcacattcaacttcactccgatcgcaaacacagcgca	844
DG		:   : : : :   : : : :   : : : :   : : : :   : : : :   : : : :   : : : :	
Db	860	twnhtnauwmaaymwumysumtmntnmntscyanmahnytuuttntnyuncscctschy	801
QY	845	aaactatttaatcagggaataactcaattggcaaggaactcctt	887
DG		:   : : : :   : : : :   : : : :   : : : :   : : : :   : : : :   : : : :	
Db	800	hcbscnmyhksammmamnmhmntscaytuymnmwscscyut	758

RESULT 12	
CNS00FOO/c	
LOCUS	1101 bp DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC: BACR31F03 of RPCI-98 library from <i>Drosophila melanogaster</i> (fruit fly), genomic survey sequence.
ACCESSION	AL070854
VERSION	AL070854.1 GI:4950896
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	<i>Drosophila melanogaster</i> Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 1101)

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RP01-98 and was constructed by partial *NotI* digestion of drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the RACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES	BASE COUNT	ORIGIN
source	167 a	84 c 86 g 155 t 609 others
Location/Qualifiers		
1. 1101		
/organism="Drosophila melanogaster"		
/db_xref="taxon:7227"		
/clone_lib="RPCI-98"		
/clone="BACR31F03"		
/note="end : 47"		

Query Match 2.5%; Score 44.2; DB 12; Length 1101;



Search completed: July 26, 2002, 03:34:28  
Job time: 2241 sec





```

/protein_id="CAD10302.1"
/db_xref="GI:16541563"
/translation="MRKISVGICITILLSLVWLOGKSSHSRSGELAINIRDEP
RSLDPROVRLLSLVSLVXHYELGIVQENLSNTEPALAEDYSLSDGLTYTFKLKSA
PWSNGDLTRADIESKWOVATQEVFSYFALNPINKVRKIQEGLSIDHFGVHSPN
ESTLVLTLESPTSHFLKLLALPVPFVHKQSRTLOSLSLP IASGAFYPRNKIQKQWIK
LSKPHYNSQVETKTIITHEIPDNTAAKLFLNOGLKNNQWGPWGERIPDPTLSNLO
SKGHLHSDVAGTWSLTFNINKFPLNNMKLREALSALDKREALVSTIFLGRKTADEL
LPTNHSYPERHQEMAQOAVAKLFLKEALEELQITAKDLEHLNLIIPVSSASSLL
VOLIREQWKSLSFALPIVKGFSALLQADLSGNFSLATGWNFAFDPMFLFIIFAY
PSGVPVVAIHHKQFLEILQNIQEQBDHQRSELVSQASLYLETFHFIIEPIYHDAFQFA
MKNKLSLNGVSPSTGVVDFFRYAKEN"
BASE COUNT      560 a 439 c 294 g 506 t
ORIGIN

```

```

Query Match      100.0%; Score 1799; DB 6; Length 1799;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1799; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acttccccctgtcaaaactatgctcagataatgctgctatgattgacaggtcttagggggag 60
DB 1 acttccccctgtcaaaactatgctcagataatgctgctatgattgacaggtcttagggggag 60

QY 61 aaaaatttcaaaaaaactcagttcctcggaatttgatgctgcaagatcagtgagg 120
DB 61 aaaaatttcaaaaaaactcagttcctcggaatttgatgctgcaagatcagtgagg 120

QY 121 aatctgtatcacctctccttagctctccgttagtctcccaaggtgcgaagagccag 180
DB 121 aatctgtatcacctctccttagctctccgttagtctcccaaggtgcgaagagccag 180

QY 181 tcactctctacatctcogggagaactcgcgtattataataagagatgaaccccggtcttt 240
DB 181 tcactctctacatctcogggagaactcgcgtattataataagagatgaaccccggtcttt 240

QY 241 agatccaagaagaagtgagactcttccgaataacagccttctcaaaatctctatgagg 300
DB 241 agatccaagaagaagtgagactcttccgaataacagccttctcaaaatctctatgagg 300

QY 301 attagtcgaagaaataatcttcaggaataatagagcctgctcttcgagaagactac 360
DB 301 attagtcgaagaaataatcttcaggaataatagagcctgctcttcgagaagactac 360

QY 361 tcttctcoggaaggaactcactatctttaaactgaatcagctcttttggagtaattgg 420
DB 361 tcttctcoggaaggaactcactatctttaaactgaatcagctcttttggagtaattgg 420

QY 421 cgacctctaaagctgaagactttatagaatcttggaaacagactactcaagaagt 480
DB 421 cgacctctaaagctgaagactttatagaatcttggaaacagactactcaagaagt 480

QY 481 ctcaagaatctatgcttttgccttgatcccaattaaaaatgacgaagaatccagaagg 540
DB 481 ctcaagaatctatgcttttgccttgatcccaattaaaaatgacgaagaatccagaagg 540

QY 541 acactctccatagaacatttggagtgcactctcctctaataatgaatcctacactgtttac 600
DB 541 acactctccatagaacatttggagtgcactctcctctaataatgaatcctacactgtttac 600

QY 601 cctggaatcccccaactcgcgtattctctaaactttttagctcttccagctcttttccccgt 660
DB 601 cctggaatcccccaactcgcgtattctctaaactttttagctcttccagctcttttccccgt 660

QY 661 tcataaatctcaagaagaccctgcaatcccaaatctctacatagcgaagcgagctttcta 720
DB 661 tcataaatctcaagaagaccctgcaatcccaaatctctacatagcgaagcgagctttcta 720

QY 721 tcttaaaaaatacaaaaaaacaatggataaaactctcaaaaaacccctcactactataa 780
DB 721 tcttaaaaaatacaaaaaaacaatggataaaactctcaaaaaacccctcactactataa 780

QY 781 tcaagtcaggtgggaactaaaaacgattacgattacattccctcccgatgcgaacacagc 840

```

RESULT 2  
 AE001606  
 LOCUS  
 DEFINITION  
 Chlamydia pneumoniae section 22 of 103 of the complete genome.  
 AE001606  
 AE001606 AE001363

```

DB 781 TCARAAGTCAGTGGAAACATAAACGATTACGATTCACTTCATCCCGATGCAAAACACAGC 840
QY 841 agcaaaactatttaatacagggaaaaactcaattggcaaggacctctcttggggagaaacgcat 900
DB 841 AGCAAAACTATTTAATCAGGAAAACTCAATTGGCAAGGACCTCTCTTGGGAGAACGAT 900
QY 901 tctcaagaaaacctctccaaatttacagctcaaggggcaacttacctcttttgatgtcgc 960
DB 901 TCCTCAAGAAACCTATCCAAATTTACAGCTTAAGGGCACCTACACTCTTTTGATGTGCG 960
QY 961 aggaacctcagtgctcacctccaatcaatcaataaattccccctcaacaatatgaagcttag 1020
DB 961 AGGAACCTCATGCTCAGCTCAGCTCAATATCAATTAATTCCTCCCTCAACATATGAAGCTTAG 1020
QY 1021 agaagccttagcatcagccttagataagaaagctcttgcctcaactatattcttaggcgc 1080
DB 1021 AGAAGCCTTAGCATCAGCCTTAGATAAGGAAGCTCTTTGTCTCACTATATATTCTTAGGCCG 1080
QY 1081 tgcataaactgcgacatcatctcctcaactcaaaattattcagctatccccgaacataaaa 1140
DB 1081 TGCATAAACTGCGCATCATCTCTTACCTTACAAATATTTCATAGCTATCCCGCAACATCAAAA 1140
QY 1141 acaagagatggcacaacgcgaagcttacgctcaaaaactctttaaagaagctttagaaga 1200
DB 1141 ACAAGATGGCACACGCCCAAGCTTAGCTAGCTTAAAAAACCTCTTTAAAGAAGCTTTAGAAGA 1200
QY 1201 actccaaactcgtctaaagatctctgaacatcttataatcttctcccttctccgttctcgtc 1260
DB 1201 ACTCCAAATCAGCTGCTTAAAGATCTCGAATCTTAATCTTATCTTTCCCGTTCTCTCGTC 1260
QY 1261 agcaagctcttactagtcacacttatcagacacagctggaaagaagcttttaggttcgc 1320
DB 1261 AGCAAGCTCTTACTAGTCCCACTTATACGAGACAGTGGAAAGAAAGTTTAGGGTTCCG 1320
QY 1321 tatccctattgtcggaaaggaaatttgcctctctcccaagcagacctatcttccagggaactt 1380
DB 1321 TATCCCTATTGTGCGAAAGAAATTTGCTCTCTTCTCCAAAGCAGACCTATCTTCAGGGAACCT 1380
QY 1381 cctcttagctcaaggaggtggttcgcagacttgcgtactatggaatttcttaacgat 1440
DB 1381 CTCCTTAGCTACAGGAGGATGGTTGCGAGACTTTTGCTGATCTTATGSCATTTCTTAACGAT 1440
QY 1441 ctttgcctattccatcaggttctcctctatgcaatcaaccataagacctctcctagaat 1500
DB 1441 CTTTGCTTATCCATCAGGAGTTCCTCTCTTATGCAATCAACCATAGAGACTTCCTTAGAAAT 1500
QY 1501 tctcaaaaacatagacaagagcgaagatcaccaaaacgcctcggaattagtgctcgaagc 1560
DB 1501 TCTACAAAACATAGACAAGAGCAAGATCACCAAAACGCTCGGAATTAGTGTCGAAGC 1560
QY 1561 tctcttaccctagagaccttctcattattgagccgacttaccacagcgcattcaatt 1620
DB 1561 TTCTCTTTACCTTAGAGACTTTCTATATTTATTTAGCGGATCTTACCACGACGATTTCAATT 1620
QY 1621 tgcctgaataaaaaactttctaatcagaggtctcaccacagaggttgggactctcg 1680
DB 1621 TGCTATGAATATAAATCTTCTATCTTAGAGTCTCACCACAGGAGTGTGGACTTCG 1680
QY 1681 ttatgctaaaggaaaattagcacctcttttaattctcgaaaactgtcgaagaactgaatctt 1740
DB 1681 TTATGCTAAGGAAAAATTAGCACCTCTTTTAACTCGCAAACTTGTCAAGAACTGATATCTT 1740
QY 1741 atactaaactgggtgccttctgtggcactcgtttctcttctgaactgtctctctctcta 1799
DB 1741 ATACTAAACTGGGTGCCTTTGTGGCACCTCGTTTCTCTTCTGACTGCTCTTCTCTCTA 1799

```



ATVALSGKKIDFLMRTTELESLPRLPIIILLIVIFHHGLLELLIILAMITITGWIPIS  
RIYGFOLILKKNKPVLSAKAMHASTFHILKKNHLLPNTLAPIISTLIIFTIPNAYITEA  
FISFGLGLQPQASLSAGLILVKEGINAIDYIPWLFPPSLIMIALSISFNLIIGBKATL  
CLEEGSHG"

gene

9727..10581

CDS

/gene="oppb"

/note="Cp0201"

9727..10581

/gene="oppb"

/codon\_start=1

/transl\_table=1

/product="Oligopeptide Transport ATPase"

/protein\_id="AAB18354.1"

/db\_xref="GI:4376472"

/translation="MDNYLLNKKDLITSTNPKRLIENSLQLENKRNRLALVSGSGS  
GKTTIKAILGFLPENCILKTSILFEDIDITKLSPELKHKIRGQKIATILQAMGSL  
TFSMRIGMIIETLRHKKNKKEAYNRAMOLLTDVCPNPKYFSFYSPFELSGMRQ  
RVVIALASOPKILIADEPTALDSMSOAVLRILRNIOQKOATILAVTNLSLVK  
ELCNDICILKDGKLIETGTVEEIFLSPKHPYTLKLLNAVSKIPKTSPPILKKNFQP  
LMSMOGGI"

gene

10578..11366

CDS

/gene="oppb"

/note="Cp0202"

10578..11366

/gene="oppb"

/codon\_start=1

/transl\_table=1

/product="Oligopeptide Transport ATPase"

/protein\_id="AAB18355.1"

/db\_xref="GI:4376473"

/translation="MTLLLSIKDLSLTIRGKKILNHLINMLIKSVLYTIVGPSGSGS  
SLAUIIDLLKPTGTITFHPDPKIPRAKRVQVIVDDISLNPCKMSIKGIISPELNI  
IGTYSKAEQNKIYNVLDVNLKPSVLHLPKLSGGOKORIALAKALSKPELLICD  
BPLSLDTLMSLILDLFOTIKKEYONTLLPITHDMSAAYIADIIVMDQSLVERH  
CREKIFTPKHITTDLLDAIPFISLISTEMEPSEYELQVASK"

BASE COUNT

3743 a 2799 c 1831 g 3275 t

ORIGIN

Query Match 100.0%; Score 1799; DB 1; Length 11648;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1799; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acttccccctgctaaactatgctcagataatgctgctgctatgattgctgaggtctaggggag 60  
|||||  
DB 953 ACTTCCCCCTGCTAAACTATGCTCAGATAAATGCTGCTATGATTCGAGGCTCAGGGGAG 1012  
|||||

QY 61 aaattttcaaaaaactctagatttccgggaattcgtatgctgagagatatcagtg 120  
|||||  
DB 1013 AAAATTTTCAAAAAAATCTAGTATTCCGGAAATTCGTATATGCGCAAGATATCAGTGG 1072  
|||||

QY 121 aatctgtatcaccattctcttagctctccgtagctctccgtagctcgaagctgcaagagtcag 180  
|||||  
DB 1073 AATCTGTATACCATTTCTCTTAGCCTCTCCGTAGTCTCCAGGCTGCAAGGAGTCCAG 1132  
|||||

QY 181 tcaactctctacatctcgggagaaactcgtctattaataagagatgaaccccgctcttt 240  
|||||  
DB 1133 TCATCTCTCTACATCTCGGGGAGAATCGCTATTAAATAAGAGATGAACCCCGTCTTT 1192  
|||||

QY 241 agatcccaagacaagtgcgactctcttcagaaatcagccttgcaaacatatctatgaggg 300  
|||||  
DB 1193 AGATCCAGACAAGTGGGACTTCTTTTCAGAAATCAGCCTTGTCAAAACATATCTATGAGG 1252  
|||||

QY 301 attagtccaagaaataatttttcaggaaatatagagctgctctgcaagaactactc 360  
|||||  
DB 1253 ATTAGTTCAAGAAAAATAATCTTCAGGAATATAGACCTGCTCTTSCAGAGACTACTC 1312  
|||||

QY 361 tctttctcggagagactcaactatacttttaaaactgaaatcagctcttttgagtaatgg 420  
|||||  
DB 1313 TCTTTCTCTCGAGGAGTCACTTATACATTTAAACTGAATCAGCTTTTGGAGTAATGG 1372  
|||||

QY 421 cgaaccttaacagctgagagacttatagaattcttgaaacaagtagtactcaagaagt 480  
|||||

DB 1373 CGACCCTTTAAACAGCTGAAGACTTTTAGAATCTTTGGAAACAAGTAGTACTCAAGAAGT 1432  
QY 481 ctccaggaatctatgcttttgccttgaaatcccaattaaaaatctacgaagatccaagagg 540  
|||||  
DB 1433 CTGAGGAATCTATGCTTTTGCCTTGAATCCAATTAATAATAGTACGAAGATCCAAGAGG 1492  
|||||

QY 541 acactctctccatagaccattttggagtgcaactctcctaataatgaatctacactctgtttac 600  
|||||  
DB 1493 ACACCTCTCCATAGACCAATTTGGAGTGCACCTCTCTTAATGAATCTACATCTGTGTGTAC 1552  
|||||

QY 601 cctggaatcccccaactcgcatttttaaaacttttagctcttccagctcttttcccggt 660  
|||||  
DB 1553 CTTGGAATCCCCCAACCTCGCAATTTCTTAAACATTTTACCTCTCCAGCTCTTTTCCCGGT 1612  
|||||

QY 661 tcataaatctcaagaagacctgcgaatccaaatctcctaactatagcgaagggagcttttta 720  
|||||  
DB 1613 TCATAAATCTCAAGAAGACCTCGCAATTCCTAAATCTCTACCTATAGCAAGGAGAGCTTTCTA 1672  
|||||

QY 721 tcttaaaaaatacaaaaaaacaatggataaaactctcaaaaaactctcaaaaaactcactataa 780  
|||||  
DB 1673 TCCTAAAAATATCAACAAAAACAATGGATTAATCTCAAAAAACCTCTCACTACTATAA 1732  
|||||

QY 781 tcaaaagtcaggtggaaactaaacgatttaacttcaacttcccgatgcaaacacagc 840  
|||||  
DB 1733 TCAAAAGTCAGGTGGAACTAAACAGATTACGATTCACTTCAATTCCTCGATGCAAAACACAGC 1792  
|||||

QY 841 agcaaaactatttaactcagggaaaaactcaattggcaaggacctctctggggagaaagcgt 900  
|||||  
DB 1793 AGCAAAACTATTTAATCAGGGAANAATCAATGGCAAGGACCTCTCTTGGGGAAGACGAT 1852  
|||||

QY 901 tcttcagaacacctatccaattttacagcttaaggggcaactacactcttttgatgtgc 960  
|||||  
DB 1853 TCCTCAAGAAACCTATCCAATTTACAGTCTAAGGGGCACTTACACTCTTTTGTATGTCG 1912  
|||||

QY 961 aggaacctatgctcacttcaactatcaataataaattccccctcaacaatatgaagcttag 1020  
|||||  
DB 1913 AGGAACCTCATGGCTCACCTTCAATATCAATAAATTCCTCCCTCAACAATATGAAGCTTAG 1972  
|||||

QY 1021 agaagccttagctacagccttagataaggaagctcttctgctcaactatctcttagccg 1080  
|||||  
DB 1973 AGAAGCCTTAGCATCAGCCTTAGATAAGGAAGCTCTTGTCTCACTATATATCTTAGGCCG 2032  
|||||

QY 1081 tgaataactgcgactatctctcactacataatatttcactatccccgaacatcaaaa 1140  
|||||  
DB 2033 TGCAAAAACCTGCCGATCATCTCTACCTACATAATATTCATAGTATCCCGAACAATCAAAA 2092  
|||||

QY 1141 acaagagatggcaacaacccaagcttagctgtaaaaaactctttaagaagcttagaaga 1200  
|||||  
DB 2093 ACANGAGATGGCAACACGCCAAGCTTAGCTTAAAGAACTCTTTTAAAGAAAGCTTTAGAAGA 2152  
|||||

QY 1201 actccaaatcactgctaaagatctcgaaactcttaattcttcttctccgcttctcctc 1260  
|||||  
DB 2153 ACTCCAAATCACTGTAAAGATCTCGAACATCTTAATCTATCTTTCCCGTTCCTGCTC 2212  
|||||

QY 1261 agcaagttcttactagtcacaacttatacagagacagtggaagaaaggttaggggtcgc 1320  
|||||  
DB 2213 AGCAAGTCTTTTACTAGTCCAACTTATACGAGAACAGTGGAAAGAAAGTTTAGGGTCCG 2272  
|||||

QY 1321 tatccctattgtcggaaaggaatttgccttcttcccaagcagacctatctcaaggaactt 1380  
|||||  
DB 2273 TATCCCTATTGTGCGAAAGGAATTTGCTCTTTTCCCAAGCAGACCTTCTTTCAGGGAACTT 2332  
|||||

QY 1381 ctttttagctacaggaggtgtgttcgcagactttgctgactctgactgacatttctaagct 1440  
|||||  
DB 2333 CTCTTTAGCTACAGGAGGATGGTTCGACACTTTGCTGATCTTATGCAATTTCTAAGCAT 2392  
|||||

QY 1441 ctttgcttatacgaaggttctcctcttatgcaatcaacccaatgaagacttccctagaat 1500  
|||||  
DB 2393 CTTTTCCTTATCCCATCAGGAGTCTCTCTCTTATGCAATCAACCAATAGGACTTCTCTGAAT 2452  
|||||

QY 1501 tctcaaaaaataagacaagagcaagatacccaaaaaacgctcggaattagtgctgaagc 1560  
|||||  
DB 2453 TCTACAAAACATAGAACAGAGAGAGATCACCAAAACGCTCGGAATTAGTGTGCAAGC 2512  
|||||

```
QY 1561 ttcttctaactagagaccttttcattattattgagcgatctaccacgacgcatcattcaatt 1620
|||||
Db 2513 TTCTCTTTACTAGACCTTTCATATTATTGAGCGGATCTACCACGACGCAATTCATTT 2572
|||||
QY 1621 tgcataaataaaactttctaacttagagagctcaccacaagaggttgagacttcg 1680
|||||
Db 2573 TGCTATGAATAAAACCTTTCTAATCTAGGAGTCTCACCACAGGAGTTGGACTTCCG 2632
|||||
QY 1681 ttatgtaagaataatagacacttttttaactctgcaaacctgtcagaactgaatttt 1740
|||||
Db 2633 TTATGCTAAGAAATAGACCTCTTTTATCTCGCAACCTGTCAAGAACTGAATCTT 2692
|||||
QY 1741 ataataactgggtgctttgtggaactcogtttcctttctgactctctctctctcta 1799
|||||
Db 2693 ATACTAACTGGTGTCTTGTGGCAGCTGCTTCTCTTCTGCTCTCTCTCTCTA 2751
|||||

RESULT 3
AE002216/c
LOCUS
DEFINITION
Chlamydomophila pneumoniae AR39, section 44 of 94 of the complete
genome.
ACCESSION
AE002216
VERSION
AE002216.2
KEYWORDS
GT:8163460
SOURCE
Chlamydomophila pneumoniae AR39.
Chlamydomophila pneumoniae AR39
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
REFERENCE
1 (bases 1 to 11764)
Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39
Nucleic Acids Res. 28 (6), 1397-1406 (2000)
JOURNAL
20150255
MEDLINE
10684935
PUBMED
2 (bases 1 to 11764)
Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Direct Submission
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On Jun 1, 2000 this sequence version replaced gi:7189484.
LOCATION/Qualifiers
1. 11764
/organism="Chlamydomophila pneumoniae AR39"
/strain="AR39"
/db_xref="taxon:115711"
/note="synonym: Chlamydia pneumoniae AR39"
complement(109. 1707)
/gene="CP0572"
complement(109. 1707)
/gene="CP0572"
/note="similar to GB:M57689 SP:P24141 GB:X56347 PID:143603
PID:40005; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="peptide ABC transporter, periplasmic
peptide-binding protein, putative"
/protein_id="AAF38391.1"
/db_xref="GI:7189485"
/translation="MRKISVGICITLLSLVSVLQCKSSHSSTSRGELAINRDRP
RSLDPQVRLSEISLVKHIEGLVQVNNLSGNIPEALAEISLSDGLTITFKLSA
FWSNGDPLAEIEFWSKVATQEVSVIAPALNPKVRIQEGHLSIDHFGVSPN
ESTLVLTSPSTSHFKLLALPFFPVFKSQRTLQSKSLPIASGAFYKPKIKQKOWIK
LSKNPHYNQSOVETKTIITHIPDANTAALFNQCKLWQGNQPPWGERIPQETLSNLQ
```

gene

CDS

```
SKGHLHSFDVACTSWLTFNINKFPLNNMKLREALASALDXKRALYSTIFLGRAKTADHL
LPNTNHSYEHQKQRAYAKLKFALBELQITARDLEHLNLIFPVSSASSLL
VOLIRQNKSEIGFAIPVIGKEFALLQADLSSCNFSLATGGWGFADFPWAFITIFAY
PSGVPYALNHRKDFLEILQNTQEQDQHQKSELVSCASLYLLETHIIEPIYHDAQFA
MNKLSNLGVSTQVWDFRANK"
complement(1653. 2687)
/gene="CP0573"
complement(1653. 2687)
/gene="CP0573"
/note="O-sialoglycoprotein endopeptidase; identified by
match to PFAM protein family HMM PF00814"
/codon_start=1
/transl_table=1
/product="O-sialoglycoprotein endopeptidase"
/protein_id="AAF73688.1"
/db_xref="GI:8163461"
/translation="MLPLGLESSCDETAIYNEDKQILANIASQDIHASYGVVPE
LASRAHLHIFQVINKALQNNLLIEDMDLIATQTPLGLSLGVSQVGHFKGIAIGAK
KSLIGVNHVEAHLHYAAYMAQNVQFVALGLVYSGAHTAAFTIENPTSYKLGKTRDRA
IGTFDKVGRFELGLPYAGAPLIEKLALEGESEDSYFSPAKVPNYDFSFGLKTVLYA
IKGNSSPRSPAPEISLEKORDIAASFQAACTTIAOKLPTTIKERSCSILIGGVA
INEYRSATQACNLPIVFFPAKLCSDNAAMIAGLGGENFQKNSIPEIRICAYQWE
SYSPSLASP"
2771. 3214
/gene="CP0574"
2771. 3214
/gene="CP0574"
/note="similar to GB:M17532 SP:PI5282 GB:X13968 PID:145356
PID:43307; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="arginine repressor"
/protein_id="AAF38392.1"
/db_xref="GI:7189486"
/translation="MKKKVVIDEALKEILRLGCAATQELCAKLAGGATTOSSVSR
WLKIQAVKAVAGRGARYSPSTETTRHLVLSIRHNASLIVITVPQASWIAAL
LQGLKDELGTLAGDDTIFVTIDEGRLPLLMVSIANLLQVFLD"
3263. 3916
/gene="CP0575"
3463. 3916
/gene="CP0575"
/note="similar to SP:PI0345 PID:41572 GB:U00096
PID:1651370 PID:1787030; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="amino acid ABC transporter, permease protein"
/protein_id="AAF38393.1"
/db_xref="GI:7189487"
/translation="MDHWLAIAIRLLRGCGYTLVSGIGILGCGSILGILGTVTSLYF
PSKLTLLANSYTVIRGTPLEIQLIILYFGLPEVLPIEPTPLVAGIILSMNSAYL
AENIRGGINSISIGOWESAMVLGKYQIFVYIYPOVFNILPSLTNEFVSLIKESS
ILMVGVPELTKTVDIVSRLEPNMENVLICAGLYFLMTTTSFSCISRLSEKRSYDN"
3906. 4586
/gene="CP0576"
3906. 4586
/gene="CP0576"
/note="similar to SP:P41076 PID:624632 GB:U00096
PID:1651271 PID:1778570; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="amino acid ABC transporter, ATP-binding protein"
/protein_id="AAF38394.1"
/db_xref="GI:7189488"
/translation="MTIRVRNLAYSVNKKKILDCGVTSLEGRHTLFLVCKSCSKMTI
LRLAGLVQPTQGDVIEGEAPALVQPELFSWMTVLGKCTHPQIHKGASTBEARE
KAFELHLLDIEEVAKNIPDLGGGQKQKQVAIVRSLCMDKHTLLDFDEPISALDPATA
SERHLELTQDELTVGLTTHDMQKVFVSCLDRIYLIQDQTVAGVYDKRDELDGSHPL
SKYIHSAQ"
complement(4620. 5978)
/gene="CP0577"
complement(4620. 5978)
```

gene

CDS

```

/genes="Cp0577"
/notes="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF38395.1"
/db_xref="GI:7189489"
/translation="MKRRHSFDSTTKKRAVSKAQIKLIKIMETDPSLNVETPNABE
ESTLOEIKRKOLESKOAEDGILLEYKQCSQFLNENWNASLALSIGSVIELASLK
OLVERIESLGGQDQDLOIOLVIEISDKPLSIGETLSGNLDMNONVIGLLIKENPE
KEMASVGYOVLLEPLSKRIGETHAKVATHDVNISSLOFHMMSVAGGRPRGHIDNNG
YRVGLGPKNGEDAVSKDYLERVSSQLTIDKVEDKPITKPNKGLLYSQSTSKPLE
GELPLGLTSGISGFTWKSASGNDGFPFSAHRKHETSDCFOTITSITSGNAG
TYTWSLKVLPVIFQIEKPEVOLSLVSYEDWLPIDNIFNMSQRPETIPALAGQTM
LAGQYDILELAHQTNQTLMSPCNSRSLQKQTNFENSPVDFYVHAHSCHWS
GP"
complement(6069..9497)
/genes="Cp0578"
complement(6069..9497)
/genes="Cp0578"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF38396.1"
/db_xref="GI:7189490"
/translation="MKRRSLKILGICLSSIVLGLFIPOLLSTESKYLIVFLSIH
KESGISCSAEELKISWFGROTARKIKLGEARDEVFSAEKFELGSLRLILYKKPKG
ITLSLSEKIMEPASIDHPSVSHLDPSLLTYLNDCKIIEHGFITMKVSSGSLVS
WHLLESYEDINLSAEATHINDCKISMTASGEGNIOKMLGHKHTFVIVGSSSF
IEKPELASALNCQIIPLSPTISQIHATVSYAKIPDKITKWHIEITISQALEPVA
HPKPDNALALORLKLIGKTEKESDIYASSVVLGASPSHLNGLISIDNKKHLTK
FLQQAOLPHYLYRAIPPOPVINPLDVAIYSNIEGTIRNAHLEADAILNDPLKLG
SCSMGAKNLFKQGTQHFNKQOELISPHESFYAEARFSGKAQITDLPFPFSG
KITARENELLHAFGSPNEPIKPTTILIHGFQCSLPLSVSNHLPFLHKLTFIS
FHTDGLVTKGNIQALINPOYPOLNTRILIPLLISDESSTSPSKDLKIQGSG
EFLSLDMSITKTKQVRSVPFGSGDLNFINVYNPKDQNKLTLLSPKSEALLGE
LKLVMDFMSLSTQCTQLOWESPERSVFFKNASCPTCLLHRTANVRLDILKSC
PRETKGLSCLTLAAGLGSLSEALPLFYDNVSKETFIINDFKSRRANNIDAKIEY
DLKSCIAPRODSTLAERFSLGQVDHLFSPESREFKQTAWIHIIPSPAGIIMPSP
GLKAGISLAGPRINYSIKNAFRFGFVDINDSENLOAQIPLINSEKSIILRENLT
AHLSTNEVDKALFOENFLLAGAYSOYPTVIDKONEFLPIRYSPEBRIOQAT
LDMGKISITANTGMALFQFLDIDQKQVSVFTPTFFSVQKSGIICKRXDALIDR
IRLAWKQTDIHDRFLMTLGDPEVKKYFHNPLSKTKNFFLIRKISGSSPEVDWS
SAYARIALKSYLSGNPFSSLADKLFLSSLDGSTPPPTVHFFPWKSNFSDIENK"
complement(9494..10786)
/genes="Cp0579"
complement(9494..10786)
/genes="Cp0579"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF38397.1"
/db_xref="GI:7189491"
/translation="MFRKLPFSKKTKGKORLNNGLLQALIQSIKVLLINEAXKEA
CVLSYGLLTCVPLVFLFLSOLHFLTNLNNKWLIIKFPDYKKPIVAIVEAAYHAYE
SNIGLVUGSFFVFCAGILMLLSLEDGLNKIFRTSWTPISLKRLVSFVITIVSPMI
FIIVCGSWIYIOTIMPIQAKFLSLSHMTALYFISRFVPLLYLLYALFCYAFLPV
AIQTSALISLIGWIVFQAFPSLOYFISNFTYAGLVALPSPLLYLYTM
YLFGGALFTFIQNRGCTFIIGDKILPSCYLQITSTYILALTTRQNEGLSPLTAQF
IAKOSKPIGVSQCLDLEKEGELFFPNNGYQPVNFSELTIKIDIAKLLHREIFKK
FNPLDGIPTFENSQNFQNASKNENLTISEARRIK"
complement(10849..11649)
/genes="Cp0580"
complement(10849..11649)
/genes="Cp0580"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"

```

/codon\_start=1

Query Match 99.9%; Score 1797.4; DB 1; Length 11764;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1798; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	acttccccctgctaataactatgctcagataaagtctgcttatgattgcagggctcagggagag 60
Db	1907	ACTTCCCCCCTGCTAAACTATGCTCAGATAAATGCTGCTATGATTCGAGGTCTATGGGGAG 1748
QY	61	aaattttcaaaaaactctagattccggaaattcgatatgacgaagatatcagtgag 120
Db	1747	AAAATTTTCAAAAAACTCTAGATTTCGGGAATTCGTTATATGCGCAAGATATACAGTGGG 1688
QY	121	aatctgtatccaccattctctttagcctctccgtagtctccaaagctgcgaagagtcag 180
Db	1687	AATCTGTATCACCATTCTCTTAGCCTCTCCGTAGTCTCCAGGCTGCAAGGAGTCCAG 1628
QY	181	tcactcctctacatctcggggaacctcgctatttaataagagatgaaccccgctttt 240
Db	1627	TCACCTCTCTACATCTCGGGAGAACTCGCTATTAAATATAAGAGATGAACCCGCTTTT 1568
QY	241	agatccagacaagtgcgacttcttccagaaatcagccttgcataacatatctatgaggg 300
Db	1567	AGATCCAGACAAGTGCAGCTCTCTTTCAGAAATCAGCCTTGTCAACATATCTATGAGGG 1508
QY	301	attagtccaagaaaaataatcttcaggaaatatagagccttcttgcagaagatactc 360
Db	1507	ATTAGTTCAAGAAAAAATACTTTTCAGAAAAATATAGAGCTGCTCTTTCAGAAAGACTACTC 1448
QY	361	tcttctcgcagcagctcaacttacttcttaactgaaatcagccttttggagtaaggg 420
Db	1447	TCCTTCCCGACGAGCTCACTTACTTCTTAACTTTAACTGAAATCACTTTTGGAGTAATGG 1388
QY	421	cgaccccttaacagctgaagactttatagaactcttggaaacaagtagctactcaagaag 480
Db	1387	CGAACCCCTTAACAGCTGAAGACTTTTATAGAACTCTTGAACAACAGTAGCTACTTCAAGAAGT 1328
QY	481	ctcagaatactatgcttttgccttgaatccaatataaaatgtacgaagatccaaagagg 540
Db	1327	CTCAGGAATCTATGCTTTTGGCTTGAATCCAATTAATAATGTAAGAAAGTCAAGAGAGG 1268
QY	541	acactctccatagacacttttggagcactctcctaatgaactacacttctgtttac 600
Db	1267	ACACCTCTCCATAGACCACTTTGGAGTGCACCTCCCTAAATGAATCTACACTTTGTTGTATC 1208
QY	601	cctggaatccccacacctcgacttcttcttaaaacttttagctcttccagctttttcccg 660
Db	1207	CCTGGAATCCCCAACCTCGCACTTTCTTAAACTTTTAGCTCTTCCAGTCTTTTCCCGCT 1148
QY	661	tcataaatctcaagaacctcgaatccaaatctctacattcccgatgcaaacacagc 840
Db	1147	TCATAAATCTCAAGAACCTCGCAATCCAATCTCTACCTATAGCAAGCGGAGCTTCTA 1088
QY	721	tcctaaaaatatacaaaaaaagaatgataaaactctcaaaaaacccctcaactataa 780
Db	1087	TCCTAAAAATATCAACAAAAACAAATGGATAAACTCTCAAAAAACCCCTCACTACTATAA 1028
QY	781	tcaagtcagttgggaactaaacgatttagcttcaacttcccgatgcaaacacagc 840
Db	1027	TCAAGTCAGTGGAACTAAACCAATACGATTACATTTCCTTCCCGATGCAACACAGC 968
QY	841	agcaaaactatttaacaggggaactcaattggcgaagacacctcttggggagaacgcat 900
Db	967	AGCAAAAATAATTTAATACGGGAAAACTCAATTGGCAAGGACCTCTCTGGGGAAGACCAT 908
QY	901	tcctcaagaaacctatccaatttacagcttaagggggaactaacactctttttagtcgc 960
Db	907	TCCTCAGAAACCCCTATCCAAATTTACAGTCTAAGGGGCACTTACACTCTTTTGTATGTCG 848
QY	961	aggaacctcagctcactcaattcaataataataattccccctcaacaatatgaagcttag 1020
Db	847	AGGAACCTCATGGCTCACCCTTCANATATCAATAAATTTCCCCCTCAACAAATATGAAGCTTAG 788

Qy 1021 agaagccttagcatcagccttagataaggaagcttcttctcaactatattcttagccg 1080  
 Db 787 AGAAGCCTTAGCATCAGCCTTAGATAAGGAAGCTTCTCTCAACTATATTCTTAGCCG 728  
 Qy 1081 tgcataaactgcgcgacatccctaccatacaataatcctcgaactatccgaacatacaaa 1140  
 Db 727 TGCAAAACTGCGGATCATCTCTACCTACCAATATTCATAGTATCCGGAATCAAA 668  
 Qy 1141 acaagagatggcaacagcccaagcttacgctataaaactctttaaagaagctttaaga 1200  
 Db 667 ACAAGAGATGGCACAAACCAAGCTTACGCTAAAACCTCTTAAAGAACTTTAGAGA 608  
 Qy 1201 acccaaatcaactgcataaagatctcaaacatcttaattattcttccgcttccctcgtc 1260  
 Db 607 ACTCCAAATCACCGTAAAGATCTCGAATCATCTTATCTTTCCTGTTTCCCTCGTC 548  
 Qy 1261 agcaagcttcttactagtcacacttatagagacagcagtggaagaagcttttagggttcg 1320  
 Db 547 AGCAAGTCTTTACTAGTCCCACTTATACGAGACAGTGGAAAGAACTTAGGGTTCG 488  
 Qy 1321 tatccctattgctgaagaagatttgcctctcttcccaagcaactatcttcagggaactt 1380  
 Db 487 TATCCCTATGTGCGAAAGAAATTTGCTCTCTCCAAAGCAGACCTATCTTCAGGAACTT 428  
 Qy 1381 ctctttagctacagagagatggttcgcagactttgctgactcctatgcatttcttaacat 1440  
 Db 427 CTCTTTAGCTACAGGAGGATGTTCCGACACTTTCGTATCTTATGCGATTTCTAACGAT 368  
 Qy 1441 ctttgccttaccatcaggaagttctctcttcttgcatacaccataaggaacttctagaat 1500  
 Db 367 CTTTGCCTTATCCATCAGGAGTCTCTCTTATGCAATCAACCATTAAGGACTTCTAGAAAT 308  
 Qy 1501 tctcaaaaactatacaacagacacatcaccaaaaacgctcggaacttagtgcgaagc 1560  
 Db 307 TCTACAAACATAGACACAGACAGACATCACCAAAAGCTCGGAATTAGTGTGCGAAGC 248  
 Qy 1561 ttctcttaccatagagacotttccatattattgagccgactcaccacagacatttcaatt 1620  
 Db 247 TTCTCTTTACCTAGAGACCTTTCTATATTATGAGCGGATCTACCAACGCGCATTTCAAT 188  
 Qy 1621 tgcctatgaataaaaactttctaatctagagctctcaccacaggaagttgtggacttcg 1680  
 Db 187 TGCTATGAATAAAAACCTTTCTAATCTAGSAGTCTCACCACAGGAGTGTGCGACTTCG 128  
 Qy 1681 ttatgctaaggaataatagcactcttcttaactcgcgaactgtcacaagaactgaattt 1740  
 Db 127 TTATGCTAAGGAATTAGCACCTCTTTTATCTCGCAACTTGTCAAGAACTGAATCTT 68  
 Qy 1741 atactaaactgggtgccttcttgccactctgttctctctgactgctctctctctcta 1799  
 Db 67 ATACTAAACTGGGTGCTTGTGGCACTCGTTCCTTCTGACTGCTCTCTCTCTCTA 9

RESULT 4  
 LOCUS AP002545  
 DEFINITION Chlamydomophila pneumoniae J138 genomic DNA, linear BCT 08-DEC-2000 section 1/4.  
 ACCESSION AP002545 AB033780 AB033792 AB033793 AB033794 AB033795  
 AB033796 AB033797 AB033798 AB033799 AB036071 AB036072 AB036073  
 AB036074 AB036075 AB036076 AB036077 AB036078 AB036079 AB036080  
 AB038347 BA000008  
 VERSION AP002545.2 GI:9956082  
 KEYWORDS  
 SOURCE Chlamydomophila pneumoniae J138 (strain:J138) DNA.  
 ORGANISM Chlamydomophila pneumoniae J138  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.  
 REFERENCE 1 (sites)  
 AUTHORS Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K., Shiba, T., Ishii, K., Hattori, M., Kuhara, S. and Nakazawa, T.  
 TITLE Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA

JOURNAL Nucleic Acids Res. 28 (12), 2311-2314 (2000)  
 MEDLINE 20330349  
 REFERENCE 2 (bases 1 to 299650)  
 AUTHORS Shirai, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-JUL-2000) Mutsunori Shirai, Yamaguchi University School of Medicine, Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi 755-8505, Japan (E-mail: mshirai@epo.cc.yamaguchi-u.ac.jp, Tel: 81-836-22-2227, Fax: 81-836-22-2415)  
 COMMENT On Aug 31, 2000 this sequence version replaced gi:6172286 gi:6172288 gi:6172310 gi:6172312 gi:6172314 gi:6172316 gi:6172318 gi:6172320 gi:6172322 gi:6172324 gi:6635158 gi:6635160 gi:6635162 gi:6635164 gi:6635166 gi:6635168 gi:6635170 gi:6635172 gi:8547422 gi:8547426 gi:8547430 gi:8978373  
 AB033780-AB033781, AB033792-AB033799: Submitted (25-Oct-1999)  
 AB038345-AB038347: Submitted (14-Feb-2000)  
 AB036071-AB036078: Submitted (18-Dec-2000).  
 FEATURES  
 source  
 1..299650  
 /organism="Chlamydomophila pneumoniae J138"  
 /strain="J138"  
 /db\_xref="taxon:138677"  
 /note="Synonym: Chlamydia pneumoniae (strain J138)"  
 complement(1..282)  
 /gene="CpJ0001"  
 complement(1..282)  
 /gene="CpJ0001"  
 /codon\_start=1  
 /transl\_table=11  
 /product="CT001 hypothetical protein"  
 /protein\_id="BAA98211.1"  
 /db\_xref="GI:8978374"  
 /translation="MLGKIIRGLSLIVLICALNVGLIGITHNKLNIKALCGGVSTP  
 ATQTYIIIGIAGVICLLSFCPCFKSRHSHGSCSSGCGCHSHSDKN"  
 573..878  
 /gene="gatC"  
 573..878  
 /gene="gatC"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Glu-tRNA Gln amidotransferase (C subunit)"  
 /protein\_id="BAA98212.1"  
 /db\_xref="GI:8978375"  
 /translation="MEQPHLDREEILLAKASALQSEELIQEYQTSLSAVITSMKCA  
 LAIDIDADSCESLFMHVYVNVEDLREDSVTSDFNRBEFLRNVPESLGLVKVPAVK"  
 895..2373  
 /gene="gata"  
 895..2373  
 /gene="gata"  
 /transl\_table=11  
 /product="Glu tRNA Gln amidotransferase (A subunit)"  
 /protein\_id="BAA98213.1"  
 /db\_xref="GI:8978376"  
 /translation="MYRISALEKAVTLGELTATGVTQHFHRIIEEAGQVAFISL  
 CKEQALDELDKKRSEPEGLKAGVPVGIKDNIVHTGLKTCASRVLENYQPPD  
 ATVERIKKEDGIILKLMDFAGSTTLYSAPHTNPDLRSVPGGSSGGSAAAV  
 SARFCVPLGSDTGGSIQPAFCGVGVGFKPSYGVSRVGLVAFASLDDIGPLANTV  
 EDVALMDVSGRDPKDATSRFFRDSFMKSLSTFVPKVIQVPRFLSGLRDDIRENF  
 FSSLAIFEGEGHLDVDELILSHAVSIYILASAEATNLARFDGVGRVYRSPQAH  
 ISQYDLRSGEGFGKVMRILLGNLYLSAEARONVYKKAIVAKIVKAFRTAFKFC  
 EILAMPVCSPAFEIGELDEVTYLODIYTVAMNLAYLPALAVPSGFSKGLPLGIQ  
 IIGQGGQDQGVQGVQSFQEAQIKQLFSKRYAKSVLGGQS"  
 2370..3836  
 /gene="gatB"  
 2370..3836  
 /gene="gatB"  
 /codon\_start=1  
 /transl\_table=11  
 /product="(pet112) Glu tRNA Gln amidotransferase (B subunit)"  
 /protein\_id="BAA98214.1"



```

/db_xref="GI:8978377"
/translation="MSAAVADWESVIGLEVEHVELNTASKLFSSALLNRFCDPENTNIST
VCTGLPSGLVNLGSAVEKAVLPGCAVEGEISLRSRDRKSYFIPDSNPFOITOFEP
PLIIRGRRIKAVOGEERYELAQTHIEDAGMLAHFGEFAGVDYNRAGVPLLEIVSKP
CMFPCDAVATATSLVSLDYGISDCNMEGSIREFDVNSVRPGKSPCLRNKVEIKN
MNSFAAEALAEAKQIDVYLNQPNKMDPKLVIIPAATYRWDPKKTIVLMRLKESAE
DYKPEPDLPTQLATVETSIYRIKTLPELVDYKHYIOEYLSBDIASIILISPKNI
ATFFEVACKDCNFRSLNWNVSVEFGCKTLGVKLPSSGIFPEGVAQLVNAIDGVI
TGKIAKETADIMMESPKNPEBILKEKPELLPMSDEGELQKIIAEVVLNAPESIVDYK
NGKTKAGFLVGQIMKRTAGKAPPRVNNELLLELDKG"
4127..6895
/gene="pmp_1"
/length=1138
/gene="pmp_1"
/codon_start=1
/transl_table=1
/product="polymorphic outer membrane protein G family"
/protein_id="BAA98215.1"
/db_xref="GI:8978378"
/translation="MRFSLCGPFLVFSEFTLLSVFDTLSLSTTSLTPDSPHGDSQNA
ERSYVWAGDVYSLTGDVSTISNVDNSALNKACFNVTSGSVTFAGNHHGLYFNFISSGT
TKGAVLCCDDPQATARESGFTSLFIQSPGDIKEQGLYSKKNALLNNYVYRFEQN
QSKTKGAIAGANVTIVGNFDSYFQNAATFGGAHSSGPIQIAVNAQAEIRFACNTA
KNGSGALYSDGDDIDQWYVLFRENEALTAIGKGVACCLPTSGSSTPVPVITFS
DNKQLVFERNHISMGGAITYARKLSISSGPTLFINNISYANSQNLGAIAIDTGEI
SLSAEKTITFOGRTSLPLFNGIHLQNAKFLKQARNYSIEFYDPTTSEAGSTC
LNIINGDKNEYTGTILFSGEKSANDPRDFKSTIPONVNLGAGLYVKEGAETVSK
FTEIISPTGNAYEDLRMSOTFPLLSLEPGAGSVTVAGDPLVPSHYFGQGNKL
AATGCKVGEFFWDKINKPRPEKGNLVPNLLGNADVRSLSMOVQETHAASLIQD
RGLMDIGLFFHVSASEDNIRYHNSGGYVLSYVNEITPKHYTSMATSQLFSDKDY
AVSNREMYLGSYLYQYTTSLGNIIFYRASNPNVNGILSRFLQNLPLT:PHFLCAV
GHATNDMKTIDYANFPKNSWNNCWAIECGGSMPLLVPEGRFLFOGALIPMKQLPV
AYQDKEFTADGRFSNGSLTISVPLIGRFEKIALSDOVLVDPSFYIPDLPRKP
SCEAALVIGSDSLVPAHVSRAHVSFGTGRVHFNDYTELLORGSIECRPHARNYI
NCGSKFRF"
complement(7138..7293)
/gene="CPj0006"
/complement(7138..7293)
/gene="CPj0006"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAA98216.1"
/db_xref="GI:8978379"
/translation="MQEPLRSALLERLSEWLVLGVPSPETTRSTPEKDNALPKDSR
NRTLESL"
7515..10169
/gene="CPj0007"
/length=1138
/gene="CPj0007"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAA98217.2"
/db_xref="GI:8978380"
/translation="MPSFLVPIPLDTSSLSLSDDEGNPQSMRKLRIILAIIVLIAL
SIIILAGVVLIVLPAIGLSVSISSPAGMACALGCVMLALGIDVLLKREVPILVAS
VTTPITVPRSGISIGADSTIRSLPTYLDEGRPQSMRKLRIILAIIVLIVFIIILIA
SGVLLTVAIPGLSIISSPAEMGACALGCVMLALGIDVLLKREVPILVVPAPLEPK
VIDIDEEISRIQOPEAAEARLPEMSAFEGYIKVESHLENKMSLYDYGHPLEKPT
KHQIRVRSRLAKMPEPLDIRIFEEFEFFLSAKRLIDLIDLTLVERKILTEQLER
NNLRKAFSLYQDSIFPKLIIDNFEKAWFMILSKISICRFTIIFENHEGVAKSLHK
NAVLLERVIYRSLQVSDIGMSAKMLIHGPNFFSLEDNKKTKMHAEMLESLS
YRKFVALLSDNVVTPSPDKKWDLSGIPCDALSEISRDQWOKKHAHLKHOSLYTO
ARLDLDSKKNQENKELEAOEYIISWVRKVFETERVQERIQALQKLYPNILREE
ETTGQETVPTVQGTASSDITDILGRIEVSSREDNQESCVCVLRSHVENSWEVK
QYGPXKKEFOQKGLERFTHIELEVLQDYSKHLSFKVKNKKEVQAKFRL
KVLSDLEGLIQAESLSLTOEELPILATRALEKAVFAGSLCCALAKAPFEL
DPFQSDTQURALTFLQRAKASLEEEKRFSNLENDTAEERLLKESKOTFERAGL
GVRLAEVSTYDLRLNTWEGTPESEKVFYSWYLNLYNEEKRRKRLVEMTQYR
DFKMALEAMQFNEALLQEELISQAPSE"
10645..11358

```

```

CDS
/gene="CPj0008"
/length=11358
/gene="CPj0008"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAA98218.1"
/db_xref="GI:8978381"
/translation="MISGLFLVRRREVTVRSEETPRGVSVPSEPALEAKOEKPE
TKKILDRPLKELDQDLYIOEVFAELRKLDPKYEDRGLLTKAEKILRVDFDVVKDMM
SEFLDIORVLEEAAYVEHCQDPLENTAYEIESSOELDYACGVCGYLPFGDARADR
LKRYSKENVDRPMRTWKSWEASVMDHSYGVAELFKKAVGVLEESVAILFKFSYRD
AFVECEKAKIQDRGFKWL"
11485..12792
/gene="CPj0009"
/length=11358
/gene="CPj0009"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="BAA98219.1"
/db_xref="GI:8978382"
/translation="MWLDRYADKFIILREKKEKMEHSHLPHATWVRKASGHAKAKAA
EKERSNENOKVXDEKELSKGLAEFRNQESRRARERLRELQTLIPEVSVEVLER
QRTKYVNLNLIADIEKHYHCVEQEHYKVENKEABYRENGEKVLSAEVSECLQ
RLEDCLETWSKKLTAKESVFEMKFDATKLGKLVSDVNRLEILCEADEEMIFRIE
EEMTLRWVLEPLLFPMKNTPEKASLOYSCKEMAKVEPOCKESPTVRSSEERLERL
QDQATYVNCQERLQGSLEKSVTRCDHLEQMKHFEVQGLNFINEELLVWGAELF
TQARDLVATVPYMEFYLOYNHIREKRVSONMAKTEREIRIQATQGVWKMEDLLAED
TILKEEDYVLLDRDMLLRDKRKNRQLCNKIAAAQQRVKGF"
13104..15428
/gene="CPj0010"
/length=11358
/gene="CPj0010"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
Query Match 99.9%; Score 1797.4; DB 1; Length 299650;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1798; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 actccccccctgtaaacatgctcagaatactgctatgattgagctcagggagag 60
Db 233705 ACTTCCCCCTGCTAAATGCTCAGATATGCTGCTATGATTGCTAGGCTAGG 233764
QY 61 aaaaatttcaaaaaaactctagtagtccgggaattcgatatcgcaagatcagtg 120
Db 233765 AAAATTTCAAAAAAACTCTAGTATTCGGGAATTCGTATATGCGCAAGATATCAGTGGG 233824
QY 121 aatctgtatcacattctccttagctctcctgtagctcctccaaagctgcagagagtc 180
Db 233825 AATCTGTATCACATCTCCTTAGCCTCTCCGTAGCTCCTCCAGGCTGCAAGGAGTCCAG 233884
QY 181 tcactcctctacatctcggggagaaactcgctattataataagagatgaaccccgctct 240
Db 233885 TCACCTCTCTACTCTCGGGAGAACTCGCTATTAATATAGAGATGAACCCGCTCTTT 233944
QY 241 agatcaagaagaagtcgactctttcagaatacagccttgcacacatctatcaggg 300
Db 233945 AGATCAAGAACAAGTCGCACTCTTTTCAGAAATCAGCCTTGTCAACATATCTATCAGGG 234004
QY 301 attagttcaagaataatctttcaggaataatagagctctcttgcagagactactc 360
Db 234005 ATTAGTTCAGAAATAATCTTTTCAGAAATATAGAGCTCTCTTTCGAGAACTACTC 234064
QY 361 tcttctcgcagagactcactatctactcttttaaaactgaactcagctcttttgcagta 420
Db 234065 TCTTCTCGGAGGACTCACTATTAACCTGAAATCAGCTTTTTCGAGAGTAAAGG 234124
QY 421 cgacccttaacagctgaagactttatagaattcttggaaacaagtagctactcaaga 480
Db 234125 CGACCCCTTAACAGCTGAAGACTTTTATAGAATCTTGGAAACAAGTACTACTCAAGACT 234184

```



QY 481 ctccaggaatctatgcttttgccttgatccatccatataaataatgacgaagatcccaagagg 540  
 Db 234185 CTCAGGAATCTATGCTTTTTCCTTGAATCCAAATTAATAATGACGAAGATCCAGAGGG 234244

QY 541 acactctccatagacacatttgagtagtccactctcctaatgaatccatccacttggttac 600  
 Db 234245 ACACCTCTCCATAGACACATTTGGAGTGCCTCTCCCTAAATGAATACACTGTGTGTAC 234304

QY 601 cctggaatcccccacactcgcatttttaaaacttttagctcttccagttcttttcccggt 660  
 Db 234305 CCGTGAATCCCAACCTCGCATTTCTTAAACATTTTAGCTCTTCCAGCTTTTTCOCGT 234364

QY 661 tcataactctcaagaacctgcaatccatccatctctacotatagcaagcgagctttcta 720  
 Db 234365 TCATAAATCTCAAGAACCTCGCAATCCAAATCTCTACCTATAGCAAGCGAGCTTCTA 234424

QY 721 tcttaaaatatacaaaaaaacaatggaataaactctcaaaaacccctcactactataa 780  
 Db 234425 TCCTAAAAATATCAAAACAAAAAATGATGAATAACTCTCAAAAAACCTCTACTATAA 234484

QY 781 tcaaaagtcaggtggaactaaacgattacgattccacttcattcccgatgcaaacacagc 840  
 Db 234485 TCAAAAGTCAGGTGGAACCTTAAACGATTAGCAATTCATTCCTCCGATGCAAAACACAGC 234544

QY 841 agcaaaactatttaatacagggaactcaatctgcaaggaacctcttggggagagcagcat 900  
 Db 234545 AGCAAAATCTATTAATCAGGGAACCTCAATTTGGCAAGGACCTCTTGGGAGAACGCAT 234604

QY 901 tctcagaacacctatcccaatttacagtcgaaggcgacttaoactcttttgatgctgc 960  
 Db 234605 TCCTCAAGAAACCTATCCAAATTTACAGTCTAAGGGGCACTTTACACTCTTTGATGTCG 234664

QY 961 aggaacctcaatgctcaacttcaataataaattccccctcaacaatatgaagcttag 1020  
 Db 234665 AGNAACCTATGCTCACTCTCAATATCAATAAATTCOCCTCAACAATAGAAGCTTAG 234724

QY 1021 agaagcttagcatcagccttagaaggaagcctctgtctcaactatattttaggcg 1080  
 Db 234725 AGAAGCTTAGCATCAGCCTTAGATAGGAAGCTCTTGCTCAACTATATCTTAGCGG 234784

QY 1081 tgaaaaaactgcgcatctctcactcaaaatattcagctatccgaacatcaaaa 1140  
 Db 234785 TGAaaaaactgcgcatctctcactcaaaatattcagctatccgaacatcaaaa 234844

QY 1141 acaagagtggaacagcagccttacgctcaaaaactctttaaagaagcttttagaaga 1200  
 Db 234845 ACAAGAGATGGGCAACGCCAAGCTTACGCTAAAAAATCTTTAAAGAAGCTTTAGAAGA 234904

QY 1201 actccaaatcactgctaaagatctcgaaactcttaactcttattcttcccgcttctcgtc 1260  
 Db 234905 ACTCCAAATCACCGCTAAAGATCTCGAATCTTAATCTTATCTTTCCCGTTTCTCTGTC 234964

QY 1261 agcaagttcttactagtcacactttacagagacagtggaagaaagtttaggttcg 1320  
 Db 234965 AGCAAGTTCTTACTAGTCCAACTTATAGCAGACAGTGGAAAGAAATTTAGGGTTCG 235024

QY 1321 tatccctatgfcggaagaaatttgcctcttccaaagcagacctatcttcagggaaatt 1380  
 Db 235025 TATCCCTATTTCGGAAGGAATTTGCTCTTCTCAAGCAGACCTATCTTCAGGAACTT 235084

QY 1381 ctcttagctacagagagtggttcgcagactttgtgactctatggcatttctaacgat 1440  
 Db 235085 CTCTTAGCTACAGAGAGTGGTTCGACACTTTGCTGATCTATNGGCAATTTCTAACGAT 235144

QY 1441 ctttgccttatccatcaggagttcctccttatgcaatcaaccataaaggaacttctcagaat 1500  
 Db 235145 CTTTGCCTTATCCATCAGGAGTTCCTCTTATGCAATCAACCATAGGACTTCTTGAAT 235204

QY 1501 tctcaaaacatacaacaagagcagatcaccaaaaacgctcggaattagtcgcaagc 1560  
 Db 235205 TCTAAAAACATAGAACAGAGCAAGATCACCAAAAACGCTCGGAATTAGTCTCCACAGC 235264

QY 1561 ttctctttacatagagacctttcatattattagccgatctaccacacacatttcaatt 1620  
 Db 235265 TTCTCTTTACCTAGAGACCTTTCAATATTATTGAGCCGATCTACCACACGCAATTTCAAT 235324

QY 1621 tgcctatgaataaaaaacttttctaacttagaggtctccacacacagaggttggaactccg 1680  
 Db 235325 TGCCTATGAATAAAAAAATTTCTAATCTAGAGTCTCCACCAACAGGAGTGTGGACTCCG 235384

QY 1681 ttatctcaagaaataatagacacctttttaaactctcgaacactgtcgaacatgaattt 1740  
 Db 235385 TTATGCTAAGGAAATTAGCACCTCTTTTAAATCTCGCAAACTTGTCAAGAACTGAATCTT 235444

QY 1741 atactaaactgggtgcctttgtggcaacctgttctcttcgactgtctctctctcta 1799  
 Db 235445 ATACTAACTGGGTGCCTTTGTGGCACTCGTTTCTCTCTGACTGCTCTCTCTCTCA 235503

RESULT 5  
 AX349501  
 LOCUS AX349501 1599 bp DNA linear PAT 06-FEB-2002  
 DEFINITION Sequence 24 from Patent W00202606.  
 ACCESSION AX349501  
 VERSION AX349501.1 GI:18615357  
 KEYWORDS  
 SOURCE Chlamydomophila pneumoniae.  
 ORGANISM Chlamydomophila pneumoniae  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.  
 REFERENCE 1 (sites)  
 AUTHORS Ratti, G. and Grandi, G.  
 TITLE Immunisation against Chlamydia pneumoniae  
 JOURNAL Patent: WO 0202606-A 24 10-JAN-2002;  
 Chiron S.p.A. (It)

FEATURES  
 source  
 1. 1599  
 /organism="Chlamydomophila pneumoniae"  
 /db\_xref="taxon:83558"

BASE COUNT 509 a 389 c 261 g 440 t

Query Match 88.9%; Score 1599; DB 6; Length 1599;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 atgcgcaagatatacagtggaactctgtatcaccaactctcttagcctctcagtagctc 160  
 Db 1 ATGCGCAAGATATCACTGGGAATCTGTATCACCATTCTCTTAGCTCTCGGTAGTCTC 60

QY 161 caaggtgcgaaggtccagtcactctctacatctcggggaggaactcgctattataa 220  
 Db 61 CAAGGCTGCAAGGAGTCCAGTCACTCTCTACATCTCGGGGAGAACTCGCTATTATA 120

QY 221 agagatgaaccccgcttcttagatcccaagacaggtgcgactcttccagaatcagcct 280  
 Db 121 AGAGTGAACCCGCTTCTTTAGATCCCAAGACAAGTGCAGTCTCTTTCAGAAATCAGCCT 180

QY 281 gtcaaacatataatgagggattagttcaagaaaaataactttcaggaataatagacct 340  
 Db 181 GTCAACATATCTATGAGGGATAGTTCAGAAATAATCTTTCAGGAAATATAGACCT 240

QY 341 gctctgcagaagactactctcttcccgacgagcactcaactattttaaactgaaa 400  
 Db 241 GCTCTTGCAGAAGACTACTCTCTTCTCGGAGGACTCACTTATCTTTTAAACGAAA 300

QY 401 tcagcttttgagtaatgagaccccttaacagctgaagactttatagaactttgaaa 460  
 Db 301 TCAGCTTTTGGAGTAATGGGACCCCTTAACAGCTGAAGACTTTATAGAACTCTGAAA 360

QY 461 caagtagctactcaagaagctcaggaactatgtcttttgctgaatcccaataaaaa 520  
 Db 361 CAAGTAGCTACTCAAGAAGTCTCAGGAATCTATGCTTTTGGCTTGAATCCAAATTA 420

QY 521 gtacgaagatcccaagaggacacctctccatagaccattttggagtgactccta 580

Db 421 GTACGAAGATCCAAAGAGGACACCTCTCCATAGACCACTTTTGGAGTGCATCTCTCTAAT 480  
Qy 581 gaatctacacttggtgtacccgtggaatcccaactcgaatttttaaaacttttagct 640  
Db 481 GAATCTACACTTGTGTACCTGGAAATCCCACTCGCAATTTCTTAAACTTTTAGCT 540  
Qy 641 ctccagcttcttcccggttcaaatccaaagaacctgcaatcccaactctctact 700  
Db 541 CTTCAGTCTTTTCCCGCTTCAATAATCTCAAGAACCTCGCAATCCAAATCTCTACT 600  
Qy 701 atagcaacgagcttct 760  
Db 601 ATAGCAACGAGCTTCT 660  
Qy 761 aaaaacccctactactataatacaagtcaggtggaaactaaaacagattagattcaattc 820  
Db 661 AAAAACCTCTACTACTATATCAACTCAGTGGGAACCTTAAACGATTACGATTCACTTC 720  
Qy 821 attccgatgcaaacacagcagcaaaactatttaacagggaaacctcaattggcaagga 880  
Db 721 ATTCCGATGCAACACACAGCAGCAAACTATTTAATCAGGAAACCTCAATTTGCAAGGA 780  
Qy 881 cctctctgggaggaacgcttctcaagaacctatcccaatttaacagctcttaagggcgc 940  
Db 781 CTTCTTTGGGAGACGCTTCTCTCAAGAACCTTATCCAAATTTACAGTCTAAGGGGCAC 840  
Qy 941 ttacactctttgatgtcgcaggaacctcctcctcctcctcctcctcctcctcctcctc 1000  
Db 841 TTACACTCTTTTATGTCGAGGACCTCTATGTCCTACCTTCAATATCAATTAATTCCTC 900  
Qy 1001 ctcaacaatgaacttagaagaccttagcatcagctcagctcagctcagctcagctcagct 1060  
Db 901 CTCACCAATATGAAGCTTAGAGAGCCCTTAGCATCAGCTTAGATAGGAAAGCTCTGTC 960  
Qy 1061 tcaactatctttagcgtgcaaaaactcgcgactcctcctcctcctcctcctcctcctcct 1120  
Db 961 TCAACTATATCTTAGCGCGTGCAGGACCTCTATGTCCTACCTTCAATATCAATTAATTC 1020  
Qy 1121 agctatcccgaaacatacaaaaacagagatggcaaacgccaagcttaagcttaaaactc 1180  
Db 1021 AGCTATCCGAGACATCAAAAACAGAGATGCGACACGCGCAAGCTTACGCTTAAACACTC 1080  
Qy 1181 tttaagaagcttttagaagacctcaaatcaactgctcaagactcgaacactcttaactctt 1240  
Db 1081 TTTTAAAGACCTTTAGAGAACTTCAATCACTGCTTAAAGATCTCGAATCTTAACTCTT 1140  
Qy 1241 atcttcccggttctcgtcagcaagcttcttactagtcctcaacttatcagagaaactg 1300  
Db 1141 ATCTTTCCCGTTTCTCTCGTCAAGTCTTTACTAGTCCCACTTATACGAGACAGCTGG 1200  
Qy 1301 aaagaagtttaggttctcgtatccctattgctcgaagaagattgctcttctcccaagca 1360  
Db 1201 AAGAAGTTTAGGTTCGCTATCCCTATTTGTCGGAAGAAATTTGCTCTCTCCCAAGCA 1260  
Qy 1361 gaactatctcaggaactctctttagctcagaggaactgctcagagacttctgctgat 1420  
Db 1261 GAACCTATCTTCAGGAACCTCTCTTTAGCTACAGGAGGAGTTCGCGACACTTTGCTGAT 1320  
Qy 1421 cctatggacttctaaocatttcttcttctcctcagaggttctccttctgcaatcaac 1480  
Db 1321 CCTATGGCATTTCTAACGATCTTTGCTTATCCATCAGGAGTCTCTCTTATGAATCAAC 1380  
Qy 1481 cataaggaactccttagaattctcaaaaacatagaaacagagcaagatcccaaaaacgc 1540  
Db 1381 CATAGGACTTCTAGAAATCTACAAAACATAGAAACAGAGCAAGATCACCAAAACGC 1440  
Qy 1541 tcggaattagttcgcagcttctcttctcctcagagaaacttctctctctctctctctct 1600  
Db 1441 TCGGATTAGTTCGCAAGCTTCTCTTTACCTAGAGACCTTCTCTCTCTCTCTCTCTCTCT 1500  
Qy 1601 taccagcagcattcatttctgataaataaaacttctcctcctcctcctcctcctcctcct 1660

Db 1501 TACCACAGCAGATTCAATTTGCTATGAATAAAAAAACTTTCTAAATCTAGGAGTCTCACCA 1560  
Qy 1661 acaggagtgtgacttccgttatgcttaaggaataattag 1699  
Db 1561 ACAGGAGTTGTGGACTTCGGTTATGCTTAAGGAAATTAG 1599  
RESULT 6  
AE001293 11944 bp DNA linear BCT 30-OCT-2000  
LOCUS Chlamydia trachomatis section 20 of 87 of the complete genome.  
DEFINITION AE001293 AE001273  
ACCESSION AE001293.1 GI:3328597  
VERSION  
KEYWORDS  
SOURCE Chlamydia trachomatis.  
ORGANISM Chlamydia trachomatis  
REFERENCE 1 (bases 1 to 11944)  
AUTHORS Stephens,R.S., Kaiman,S., Lammel,C.J., Fan,J., Marathe,R.,  
Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,  
Koonin,E.V. and Davis,R.W.  
TITLE Genome sequence of an obligate intracellular pathogen of humans:  
Chlamydia trachomatis  
JOURNAL Science 282 (5389), 754-759 (1998)  
MEDLINE 99000809  
PUBMED 9784136  
REFERENCE 2 (bases 1 to 11944)  
AUTHORS Kaiman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,  
Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.  
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)  
MEDLINE 99206606  
PUBMED 10192388  
REFERENCE 3 (bases 1 to 11944)  
AUTHORS Stephens,R.S., Kaiman,S., Lammel,C.J., Fan,J., Marathe,R.,  
Aravind,L., Mitchell,W.P., Olinger,L., Tatusov,R.L., Zhao,Q.,  
Koonin,E.V. and Davis,R.W.  
TITLE Direct Submission  
JOURNAL Submitted (20-MAY-1998) Program in Infectious Diseases, University  
of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA  
FEATURES  
source  
1..11944  
/organism="Chlamydia trachomatis"  
/strain="D/UW-3/CX"  
/db\_xref="taxon:813"  
complement(105..878)  
/gene="CT192"  
complement(105..878)  
/gene="CT192"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAC67784.1"  
/db\_xref="GI:3328598"  
/translation="MFPGSVETLPSTCLISKTRKEFRVRSVGQEAASONTLSWTIR  
PRIPITIGSKEVSLALVLTAVGACAAVGGAFSCLOVLGLGVLATGILL  
AVLEFCHRRSSMEKXVILLKQDLFEPYIQEQAPLIEASTCPGPILSGPEVQ  
QBRPVLLKDLDSLHAPKTYAVGSHVVLKAGIKGRNLEEGIDTDQNFVRVD  
ELFLLGKQEMRLDGCCKVLPLPKSKSESINNLVSDC"  
/gene="tgt"  
complement(905..2023)  
/note="CT193"  
complement(905..2023)  
/gene="tgt"  
/codon\_start=1  
/transl\_table=11  
/product="Queuine tRNA Ribosyl Transferase"  
/protein\_id="AAC67785.1"  
/db\_xref="GI:3328599"  
/translation="NALREFILHQSKSRARVRIETAHYIDTFAPFVPVATNGALG  
VLDHNSIPLMFCNTYHLIVHPGAEIAAAGGLHOFIGNAPITITDGGFOIFSLAYGS  
VAEIKSGKKGKGNITIKVNDGTVHFKSVYDGRKFLSPETSVQAKDLGAILPL  
DELLPFHADPTYPHQSSQTYWEKRSIDHLNPFQISQYGVTHGGTFDPQRKLCK

```

FVEDLPFGSAGSLGKLNQIDIVEVGVTAANLSAERPHLLGIGDPSIWATVCGF
IDFSOSSYTKAARHGMLTSGPLKNNQRYSSDLNPEFGCSCLACSGQITRAYLFR
HLFKVHEPNAGIMWASIHNMHHKQKVMREIREGILNDRI"
complement(2150..3562)
/gene="mgte"
/note="CT194"
complement(2150..3562)
/gene="mgte"
/codon_start=1
/transl_table=11
/product="Mg++ Transporter (CBS Domain)"
/protein_id="AAC67786.1"
/db_xref="GI:3328600"
/transl_table="MOSKTSHLDDLCFKLEAFDTLTAGEHSODLTSIVSVYNPIDL
AYAVSCIPSDSRSTLYKNLDSIASKIAPILINDSASWAFRNLSGCEICALEOMPP
DEAIWLDDIPDRYERILDLIDVKKALKIDKOKHGRNAGRLMNEFAFLMETTV
KEVAICINRNGIDLRFLVDFKGELOGEVTDORSIIASPEMLKQIMRPIEHKVL
AUTREEDVLDVERIKAVLPVVDENEFILGAIYEDVETIEDIADETIARMAGTE
DVGHDHVVORLLRAPMLLITLCAGLSASVMAYFOKIAPTLLAMVIFRPLVNGLE
SONGVQCSITLVRSMATGTLGSRRRRTILKEMSIGLGLGVALGILGLVYCCMGCL
GIGLFGVGGVQLGVTVSGILGASLTATTLGVLSPPFFAKIGVDPALASGPVITAND
IMSVVILLITGLNVLFRK"
complement(4001..5092)
/gene="CT195"
complement(4001..5092)
/gene="CT195"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67787.1"
/db_xref="GI:3328601"
/transl_table="WVMSNLNPPAEVRLPVTAKSPFNSSAWRATQKRITGCCSLNAS
SYRPSARVLSFVGVVLVGLGICFRTAYLANSEVITTFPMFAIALILFFLGGGLAI
LFRTAGKVDYCKKIQEPISQWERNILCEKEGFIPTQEPDMVDSVLLDRGTSG
LAPVYTPNDATVLCVTSILLFIFSVNLYNIRFFIYFVILFQWRYQYQTD
IPKERVQCDIVREMTSLQAVKAPFYGAVCYLAMYLLNPLSGRWVLSASLQDM
NDVIRSGVWGIFCEKNYMEGGTSGLGQNAWYLLGCFQPVQVFLFKDGVLSGA
RPSIQSPSEKEYLASFLYGAVPGLAGF"
5316..5636
/gene="CT196"
5316..5636
/gene="CT196"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC67788.1"
/db_xref="GI:3328602"
/transl_table="NATVNVNYSLPCEKMWSSATLAAHSFFNHIKLLIPLLVGVFC
ILGLAILTGVTITPISIAASYFLSLGVTLIVGAGLCAAEKRPFSVTQSKASTLLH
ISCD"
5712..6728
/gene="gcp_1"
/note="CT197"
5712..6728
/gene="gcp_1"
/codon_start=1
/transl_table=11
/product="O-Sialoglycoprotein Endopeptidase"
/protein_id="AAC67789.1"
/db_xref="GI:3328603"
/transl_table="MLTIGLESSCDETSLVQNGKILANKIASQDIHASVGGVPEL
ASRAHLQTFPELLTAATQASGVLEDIELISVANTPGLIGALSIGVNFAGLASGLKR
PLTGVNVAHLIYACMEAPATQPPALGAIISGAHTSLFLMPDATTELLIGKTRDDAI
GETFDKAVFLCLPYPGOKLELEAREGADAFAPSPARVSGYDFSGGLKTAIVIAL
KGNNSAKAPPEVSEYQKRNIAASFQKAVFTIAQKLPIDIVIKTFSCESLIVGGGVAN
NSVFERLLNQICSLPIITFPSSQLCSDNAAMVIAGLGERLFCNRTHVSKEVIPCARYQE
SACS"
6692..8248
/gene="oppa_3"
/note="CT198"
6692..8248
/gene="oppa_3"
/codon_start=1

```

```

/transl_table=11
/product="Oligopeptide Binding Protein"
/protein_id="AAC67790.1"
/db_xref="GI:3328604"
/transl_table="MRKLSVGICLLALLATSGCKSSSNATHRSPATHVAVSVKDDP
RTFDPREVRLLSDINLIHLYEGVQVTPSPGEVFPALAESFFLSEDKKTYTFLNKKAF
WSNGDLITADHFRSNDVQNRVASYISFAFLPDVKNKSGFFAKDDHTLVINLLTP
TPHFLKLLTPVYVPSVHQHQRKEEKSLPISTGAFFLKKKDRPWLKLEKSPYYNK
DQVAQVEIGIHIIPOQTASALPNQGLDWQGLPWGHSIPQETLATINKRARPSFDI
SGTSLTFTAKKPSHSHKLRQALSIVLNKEALASIAFVKPAKHLPAHLHTYPEQPS
YKQOEAITAKSLEELTELMNTIEDLEKYPITFSATSTANSTQIAQLMDQWRSLG
ITPICGKYALLQNDLIGNTFMSIGMFAFSDPLAFLSIFSCKGVKVALQDPQF
DQILISIEFKPNQKRSALISEASLYTERQNVIEPLYHDVFHYTTNNKLSFVLHPSG
LVMRYAKNS"
8407..9348
/gene="oppB_1"
/note="CT199"
8407..9348
/gene="oppB_1"
/codon_start=1
/transl_table=11
/product="Oligopeptide Permease"
/protein_id="AAC67791.1"
/db_xref="GI:3328605"
/transl_table="MLSYIKRRLLFNLLSLVWVTLTFFIITKIPGDPENDENGNTLS
SETLALKNRYGDKPLFTQYLIYKCLLTDGEGSLIYKDRIVISIIAALPSSAIL
GLESCLSLFGSITLGIILAAFYKKSGRTIFFSSVQISVPAFVGAFOYVFAIKYS
CLPIACWGNFSHTLLPSIALAITPMATITQULICASVSAUNKKDYVLLAYAGLSPFKV
LKHILPIALFPVVISYSAFLITMTITGTFSTENLFCIPGLGKWFICISQROVPITLG
LSVFYGAFFMTLSCCDLLQAWIDPQIRYSYKERSK"
9380..10225
/gene="oppC_1"
/note="CT200"
9380..10225
/gene="oppC_1"
/codon_start=1
/transl_table=11
/product="Oligopeptide Permease"
/protein_id="AAC67792.1"
/db_xref="GI:3328606"
/transl_table="MFSSSSHTWRYIRTNKMLVGLCLTLAVLVLAATFLPYLDYE
RTFFPHALQSPSKAFPGFDNGLRCMLARTLQGLRLSLAVSATIIDVCLGLWSTL
ALASGKVADIMSRITEILFSIPRIPVILLIVFNHGLPLILAMTINGWPIARIIT
YQGLLEKNKEFVLSARALQASTFHLRKHLPLNSIGLPIITSLIFTIPNAVYTERAFIS
FLGIGIQQPVASLGLTVKGIHSLAYHPWLFIPSFPMIIVSVFMCIGEGRLTKLLE
ENTLV"
10218..11051
/gene="oppD"
/note="CT201"
10218..11051
/gene="oppD"
/codon_start=1
/transl_table=11
/product="Oligopeptide Transport ATPase"
/protein_id="AAC67793.1"
/db_xref="GI:3328607"
/transl_table="MSEDLKIDNLVSVKDSQRVLNHLSTIKRCSMALVGENSGS
GKTTVSKAVLGLPDNCTQSGRIYSDITRLSRQLQTRGRKKIATIFQNAMGTL
TPSMRYCAQVEHLAHHEDESKAEAFSKARELLESVHIESPCRLQLYPFELSGMQQ
RVSTALANPELIIADEPSTALDSISQAQVLRVLTQIHNHSTALLITHNLALYS
ELCEEMAIRYGEIIEVQGPVQELLHSPSHPYTQQLIRAIKPIPSPYLSKPEPLATTA
Y"
11066..11809

```

Query Match 24.9%; Score 448; DB 1; Length 11944;  
 Best Local Similarity 56.7%; Pred. No. le-98;  
 Matches 976; Conservative 0; Mismatches 695; Indels 51; Gaps 6;

QY 1 actccccccctgctaaactatgctcagataaagctgctgctgattgctgagtgaggag 60  
 ||||| ||| ||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 6589 ACTTCCCTCTTCCTCCAGTTAIGTTTCAGACACGCTGCAATGATAGCGGGTTAGGAGAGA 6648  
 QY 61 aaattttcaaaaaaacctctatttccc---ggaaatttcgtatatgcgcaagatcattcagt 117

Db 6649 GACTGTTTTCGAATCGACACATGTTTCTAAGGAGGTTCATTCATCGCGCAAGATATCAGT 6708  
 QY 118 ggaatctgfatccacattctcttagcctctcgttagctctcgaagctcgaaggagtc 177  
 Db 6709 GGGAACTGCTGCTCTAGCATTAGCAACTTTCGGATGTTCAAAATCTCTCTTAACGC 6768  
 QY 178 cagtcactctctacatctcgggagaaactcgtattataataagagatgaacccgttc 237  
 Db 6769 AACCCATCGGCTCCAGCTACTCACAGATTGCTGTAAACGTAAGAGATGTCCTCCAC 6828  
 QY 238 tttagatccaaagacagtgcaactctcttcagaatcagccttgcataacatctatga 297  
 Db 6829 ATTGTGATCTCGAGAGGTTCCGCTCTCTCTGATATCAATTTGATTCATCATCTCTATGA 6888  
 QY 298 gggattagttccagaataaactcttcaggaaatataagacgtctcttcgagaagacta 357  
 Db 6889 AGGATTGGTACAGAA---ACTCTCTCGAGAGTCTTCCTCGCTTTAGCGGAGAGTTT 6945  
 QY 358 ctctctctctcggaggaactcacttatacttttaaacgaatcagctttttggagtaa 417  
 Db 6946 CTTCTTATCCGAGAGATAAAAAAGCTTATCTTCAACTTGAATAAGCTTTTTCGAGCAA 7005  
 QY 418 tggcgaccccttaacagctgaagctttatagaactcttgaaacaagttagctactcaaga 477  
 Db 7006 TGGAGATCTTATACCGCTCATGATTTTGTCTCTCGAATGATGTGTTTACAAATCG 7065  
 QY 478 agtctcaggaactctatgctttgcttgaatccaatttaaaatgaagaagatccaaga 537  
 Db 7066 TGTGCTAGTATTTATTTTTCGCTTTTCCTCTATTGA----- 7104  
 QY 538 gggacacctctccatagacaaatttggagtgcaactctctaatgaatctacactttgt 597  
 Db 7105 -----CGTGAATAAGGATTTGCGATTTTTCGAAAGATGATCATACTCTTCTAT 7155  
 QY 598 taccttgaatcccaactcgcattctttaaacttttagctcttcagctttttccc 657  
 Db 7156 CAATCTCCTCACTCCAACTCCACATTTTCTAAGCTTGCTTACCTCCCGCTATTATCC 7215  
 QY 558 cgtctat---aaatcctaaagaaacctgcactccaaatctctatagcagcgagc 714  
 Db 7216 TGTGCAATCGCAGCATCAGATCGGAAAGAGAAATACTCTCCGATATCTACTGGAGC 7275  
 QY 715 ttctatctcaaaatatacaaaacaaatggataaaactctcaaaaacccctcaacta 774  
 Db 7276 TTTTCTTAAAGAGAGAAAGAACCGCAAGATGGTTAAGCTAGAGAGAGCGCTTACTA 7335  
 QY 775 ctataatcaaaagtcaggtggaaactaaaacgattacgattcaactcattcccgatgcaaa 834  
 Db 7336 CTAATAAAGACAGGATAGCTGTACAGGAGATCTGTATACATACTCTCTGATCAACA 7395  
 QY 835 cacagcagcaaaactatttaacagggaactcaattggcaaggaactcctctgggaga 894  
 Db 7396 AACTGCTCTCTGCTTTATTAACCAAGGAAGCTAGATTGGCAAGTCTCCCTTGGGACA 7455  
 QY 895 aqcatctctcaaaacccctcaattcaattcaactcgaagggcaactcaactctttga 954  
 Db 7456 TAGTATTCGCAAGAAACTTTACCCACAAACAAACAGCAGAGCTCCCGGATCAATTGA 7515  
 QY 955 tctcgaggaaacctcgtcgcacttcaatcaatcaataaattccctcaacaatatga 1014  
 Db 7516 TATATCTGCTACTCTCTGCTTACATTTAACAACCTGCCAAAAGCTTTTAGTCATCCAA 7575  
 QY 1015 gcttagaagaaccttagcatcagccttagataaaggaagctcttctcgaactatattct 1074  
 Db 7576 GCTTCGCCAAAGCTTTGAGTCTAGTTTAAACAAAGAAAGCTCTTGCTC-----CTT 7626  
 QY 1075 aggcgctgcaaaaactccgcatctctcactcaactcaaatattoatagctatccgaaca 1134  
 Db 7627 GCCTTTTGTAAACCTGCAAAACATCTCTCTCGCATTTGACACACCTTACCCAGACA 7686  
 QY 1135 tcaaaaaacagagatgscacacgccaagcttagcctaaataaactctttaaagaagcttt 1194

Db 7687 GCCTTCTTTATAAGCAACAGAGGCCACTTACTAGTAAATCTTTACTAGAAAGACCTCT 7746  
 QY 1195 aagaacatccaaactcgttaagagatctogaacatcttattcttctccogtttc 1254  
 Db 7747 GACTGAGCTTAACATGACTATTGAGGATCTAGAGAGTATCTCTTACCTTTCCGCAAC 7806  
 QY 1255 ctccgcagcaagctcttactagtcacacttatagagaacagtcggaagaagtttag 1314  
 Db 7807 GTCTACTATGACTACAGATAGCTCAGATGTTGCGGATCAGTGCAGGAAGTTTAGG 7866  
 QY 1315 gtccgtatccctattgtcggaaaggaattgtctctctccaaagcagacactatctcagg 1374  
 Db 7867 AATTACTTTCCCTATCTGCGGGAAGAAATATGCTTTGTTGCAAAACGATCTAATAGGCAA 7926  
 QY 1375 gaactctcttagctacaggggaggttctcgcagactttgtgtatctcattggaattct 1434  
 Db 7927 TACTTTCTTTATGCTATAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 7986  
 QY 1435 aacgatcttgccttccatcagggaggttctctctctctctcgaatcaaccataaggactct 1494  
 Db 7987 TTCCATTTCTCTCT---CGAAAGGAGTCAAAACCTTATGCTTTTACAAGATCTCAATTTGA 8043  
 QY 1495 agaaattctcaaaaacatagaaagagcaagatcccaaaaacgcctcggaattagtgc 1554  
 Db 8044 TCAACTGATTCTCTCTATAGAAACGGAACCAAAACCCCTCAAAACGCTCAGCTTTAAATTC 8103  
 QY 1555 gaaagctctcttccatcagacacttctctctctctctctctctctctctctctctctct 1614  
 Db 8104 CGAAGCTTCTCTATCATAGAAAGCAAAACGTCATAGAACCCCTCTATCAGACGCTGT 8163  
 QY 1615 tcaattgtctatgaataaaaaactttctaatcttaggagtcctcaacaaagaggtgtgga 1674  
 Db 8164 CCATTATACAACAATAATAACTTTCTTTGTTAGACTACATCTCTGGGCTAGTTGA 8223  
 QY 1675 ctccgtatgctaaagaaattagcactctttaaactctg 1716  
 Db 8224 TATCGGTATGCTTAAAACTCTTAAAGATGTTTTTACTCCG 8265

RESULT 7  
 AF002215/c  
 LOCUS Chlamydomydia pneumoniae AR39, section 43 of 94 of the complete genome.  
 DEFINITION  
 ACCESSION AF002215 AB002161  
 VERSION AF002215.2 GI:8163458  
 KEYWORDS  
 SOURCE Chlamydomydia pneumoniae AR39.  
 ORGANISM Chlamydomydia pneumoniae AR39  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomydia.  
 REFERENCE 1 (bases 1 to 12980)  
 AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.  
 TITLE Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR39  
 JOURNAL Nucleic Acids Res. 28 (6), 1397-1406 (2000)  
 MEDLINE 20150255  
 PUBMED 10684935  
 REFERENCE 2 (bases 1 to 12980)  
 AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA  
 COMMENT On Jun 1, 2000 this sequence version replaced gi:7189472.  
 FEATURES Location/Qualifiers  
 1. .12980

```

/organism="Chlamydomophila pneumoniae AR39"
/strain="AR39"
/db_xref="taxon:115711"
/note="synonym: Chlamydia pneumoniae AR39"
complement(157. .1609)
/gene="CP0560"
complement(157. .1609)
/gene="CP0560"
/note="sodium:sulfate symporter family protein; identified by match to PFAM protein family HMM PF00939"
/codon_start=1
/transl_table=11
/product="sodium:sulfate symporter family protein"
/protein_id="AAF3887.1"
/db_xref="GI:8163459"
/translation="MNKKRFLSLFLTAVALGLTWSFSPASINSNANOLFAIETTTI
MGIIFQVPMGAIIGISTLLLTLOGLSGRPNPAWEVLSFSAKGLIKTG
LGERIAYFVSALGSKPLGSLVGLVIDDFLAFAIRSVARAGGLIYPVTSLSDFG
SSAEKGTQDLIGSLIKVAYQSVITSAEFLITAMGNPLVAALAGHGVSLVSLWAK
AAIPGSLFLMPLIILYKLPKPKITSCAEIRSAKIRKEMGPLKKEKTILMIFEL
LVILWTFGLIGSATTAALIGSLILNLDQKDVIANWTAWETIWTGALIMMA
SFLNQLGFIPLVDSAAALVGLSMKIGFLLFIYVSHYLFASNTAHIGAMPYFI
AVSISLTGNIPIAALTAFASNLFGSLTHYSGSPALYFGSHLVVQEWWSGFSLSI
VNIVITGIGSLMNAKALGI"
complement(1653. .2408)
/gene="CP0561"
complement(1653. .2408)
/gene="CP0561"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF3880.1"
/db_xref="GI:7189473"
/translation="MLMDISHLEDIAVDGSLIPRAIEAAIVKOMQIPTVLLHLHD
ATQVRPEVNDGSLKELIETPKINPVKAAISGLVTLVGAGKIPDKVIRVPAELL
NYRLEKQPSFANDNIGACTLYPGELFYPIKSAFDGGLVDTSFISMEDVENIHEET
VESICITLCSSTELINDLEMEKWELEDFFIEP"
complement(2487. .2684)
/gene="CP0562"
complement(2487. .2684)
/gene="CP0562"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF3881.1"
/db_xref="GI:7189474"
/translation="MKKFFVCEELCILVATHRALLETPTLALSFRELKTKYVYRAK
DILQLHNYKGFITLNTSLPCS"
complement(2812. .3000)
/gene="CP0563"
complement(2812. .3000)
/gene="CP0563"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF3882.1"
/db_xref="GI:7189475"
/translation="MGYQIFRLKGLGSFSSYNPNPNPYFFDEGCFVYVESQFKA
LQDHGLQKQTEIFVNT"
complement(3012. .3809)
/gene="CP0564"
complement(3012. .3809)
/gene="CP0564"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1

```

```

/transl_table=11
/product="hypothetical protein"
/protein_id="AAF3883.1"
/db_xref="GI:7189476"
/translation="MNTYTFSPTLQKSFSLFLEKLDYSFFFGTQTLVITPTNIR
LAARKGCKVSTIEKIIKILSFILLPLVIIAFILYFLHKKDFKOFICIPKVISNEDE
ALLGSRPOAVEKAVREISPAFFSIPBKYOLIRIDTPKDDAPSILPEGLIILKLCY
DTLKQSNLFLAKREMDFIHGPHEAKLEDSICSEKDOEWMSLESKLLITHTFLKYLVS
GIEQLNPGFENGGRGYFSISIAKIHFHQHGRYGPIKSSGPIMKET"
complement(4314. .5102)
/gene="CP0565"
complement(4314. .5102)
/gene="CP0565"
/note="similar to SP:P42065 PID:677944 GB:AL009126;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, ATP-binding protein"
/protein_id="AAF3884.1"
/db_xref="GI:7189477"
/translation="MTLISIKDLSLTTRKKILNHLNMLIKGSYLTIVGSPSGKS
SLAULTDLKPTGTITTFHMDPKIPRARKVQVIVDIDSSLNPCMSIKGICSEPLNI
IGTYSKAPQNKIYNLDVNLFPKSVLHLPKYLKSGGQKQRIAAKALSKPELLICD
EPLSLDTLNGSLILDLFQTIKKEQNTLLFITHMSAAYIADTIANWDQGSUVEHA
CREKIFSTPKHTTTDOLLDAIPIFSIJSTEMEPSEEYELQVASK"
complement(5099. .5953)
/gene="CP0566"
complement(5099. .5953)
/gene="CP0566"
/note="similar to GB:M57689 SP:P24136 GB:X56347 PID:143607
PID:580898; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, ATP-binding protein"
/protein_id="AAF3885.1"
/db_xref="GI:7189478"
/translation="MDNYLLNKDLTITNPKRTLIENLSLOKKNRNLALVGRSGS
GKTTTTRAILGLPECLIKTGSILFEDIDITKLSPKELHKIRGQKATIQONAMGSL
TPSRHKGQIIEITLQROHKNKEEAYNAKMLLDDVCIPNPKYSQYFELSGMRQ
RVYTIALASQPKLILADEHTEALDMSQAQVLRILNNQQQKQATILITHTNLSLVK
ELMNDICILIKDGKLIETIGVEIFLSPKPYTKLLNNAVSKPIPKTSSPILKNKFPQ
LMSQGGGL"
complement(5946. .6800)
/gene="CP0567"
complement(5946. .6800)
/gene="CP0567"
/note="similar to GB:M57689 SP:P24139 GB:X56347 PID:143606
PID:40007; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, permease protein"
/protein_id="AAF3886.1"
/db_xref="GI:7189479"
/translation="MENSSAPSRISWKSIQNKMLVLGTLTIIIMLGAALLPWFYQ
DYETSLLKDIIVSPCSRFPTDITGRMFARTLGRILSLLIATLIDVCGVGLW
ATAISGGKKIDFLMRTTEITSLRPIIILLVIFHGLLPILIAMTTIGWIPIS
RIIYQFLLILKNKPFVLSAKAHASTPHILKHLNLTAPLISILFTIFNAYTEA
FISPLGIGIQPPQASGLIVKEGINADYTPWLPFFPSLIMIALSISNFIKGAATL
CLEEGSHG"
complement(6831. .7775)
/gene="CP0568"
complement(6831. .7775)
/gene="CP0568"
/note="similar to SP:P24138 GB:X56347 PID:580897
GB:AL009126; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="peptide ABC transporter, permease protein"
/protein_id="AAF3887.1"
/db_xref="GI:7189480"
/translation="MFSYIKNRILFNLSLWIVLITFLVMTKTIPTGDPFNDGCVNLS
EEVLOTIKRGYGLDKPYQOYTLVHSIAKLDGNSLVYKQKVNITSTAFPSAIL
GLQSLFSLSGGGIALGTIAALKKKQRRIYLGASIIQISIPAFITAILQYVAVKIP

```

```

LPIACMGSEFTHILPTLALAVTNPFAFIQLIQTYSVSAALNDYLLAYAKGLSPKRY
WIKHILPYALFTITSYSFELTTIVITGTALENICFIPGLGKWMTCSIKQDYPVALG
LSVPYGLFWLSGSLLSDISIDPOIRYAHGKKKKR"
    complement(8069..9654)
    /gene="CP0569"
    complement(8068..9654)
    /gene="CP0569"
    /note="similar to GB:M57689 SP:P24141 GB:X56347 PID:143603
PID:40005; identified by sequence similarity; putative"
    /codon_start=1
    /transl_table=11
    /product="peptide ABC transporter, periplasmic
peptide-binding protein"
    /protein_id="AAF38388.1"
    /db_xref="GI:7189481"
    /translation="MKMHLRLTKSLIPNLLFLILLTSCKSQKPQLKHLYTAMS
HLADIAPNAYLRADLAALAEGLRETDGGIALAALAESVTLSKDHWKYTFKLPR
SWSDGTPUTAYDEKSIKOLAFEEFSYHTLIGLVKNSSAIHNAAKSLETIGLOAK
DOLTLVTLEQPFPEYLTILIRPVSPPVHHTKLPRESKNGTPPESTYINGPFLVKHEH
QNVLILEKNPAYDHESYLDRVTLATIPDASTATKLFKSKSIOWGPSAFSIGNED
QKVLSEKILIYSVSTTLLINLQKONKARLKAIAHIDRKSLIRLVPVSGOEV
TLVPPNLSQLNLQETESTQTKARAYFOQAETSEKELAELSILYPTIDSSNSGIT

```

Query Match	22.8%	Score 409.8	DB 1	Length 12980	
Best Local Similarity	54.8%	Prod. No. 2.le-89			
Matches 835	Conservative	0	Mismatches 582	Indels 6	Gaps 1
QY 185	tcctctacatctcgggagaaactcgtctattaatataagagatgaaccccgcttcttagat	244			
Db 12815	TCCCAATCCTCTAAAGAATTTTGTGTAATATGAAGAAATGCCAGCTCTCTGGAT				
QY 245	ccaagacaagtcgactctttccagaatcagccctgtcaaacatatactatgaggatta	304			
Db 12755	CCTGGAATAACTCGTCTCATTCGACCAAACTTAATCGTCATCTATTAAGGACATC	12696			
QY 305	gttcaagaaaaatactttcagaaatatagagcgtctcttcgagaagactactctctt	364			
Db 12695	GTGGAAGAACATTTCCCAAAATGGAGATTAACCCAGCCCTTGCAGAACTACACATC	12636			
QY 365	tctcgagcgaactacttatacttttaactgaaatcagcttttggagtaatgcgac	424			
Db 12635	TCCGAAGACGGACTCGGTACACATTTAAATCAAAATCATCTTTGGAGTACGGAGAC	12576			
QY 425	cccttaacagctgaagactttatgaactcttgaaacaagtagtactcaagaagcttca	484			
Db 12575	CCTCTGACAGCTCAGACATTTGTCCTCTTGTGAAGAAATCTTAAAGGAAGATCGTCC	12516			
QY 485	ggaatctatgcttttgccttgaatccaatlaaaatgtacgaagatccaaagaggacac	544			
Db 12515	TCCGTATATCTCTATCGCTTTTACCTATCAAAATGCTGGGCAATCTTTGATGATACT	12456			
QY 545	ctctccataagacattttggagtgcactctcctaataatcacacttgttggtaacctg	604			
Db 12455	GAGTCTCCGAAATCTAGGAGTCCGAGCTTAGATAAGCGTCATCTCGAAATTCAGTTA	12396			
QY 605	gaatccccaacctcgcattcttctaaacttttagctcttcagctcttttccocgtcat	664			
Db 12395	GAAACTCCCGCGGCANTTCTACATTTCTTGACTCTTCATTTTTCCTCTTCAT	12336			
QY 665	saatctcaagaacocctgcaatccaactctctctctctctctctctctctctctctct	718			
Db 12335	GAAACTCTGGAAATATAGCACCTCTTTTGAGAGATGCCCATTTACCTGGGTGCTTTC	12276			
QY 719	tatcctaaaaatacaaaaaaacatggataaaactctcaaaaaacocctcactcat	778			
Db 12275	CGCCCTGTGTCCTAGAAAAAGCGCTGAGACTCCATCTCTAGAGAAACCTTATGTACCAT	12216			
QY 779	aatcaaaagtcaggtgaaactaaaacgattacagattcaacttctccogatgcaaacaca	838			
Db 12215	AATAAAGCCGTGGAACATACATAAAATTTATTACAGTTTATCTCAACGCCTAACACT	12156			
QY 839	gcagcaaaactatttaatcagggaacactcaattggcaagacctcctttggggagaaacgc	898			

Db	12155	GCAGCCATTCTATTCAACATAAGAAATTAGATTGGCAAGGACCTCTCTTGGGGAGAACTT	120956
QY	899	attctcaagaaacctatccaatttacagctcaaggaggcaattacaaccttttggatgc	958
Db	12095	ATCCCTCCAGAAATCTCAGCTTCTTACATCAAGATGACCAGCTCTTTTCTCTTCCTGGGG	12036
QY	959	gcaggaaacctctggctcacctccaattcaataaaattccccctcaacaattatgaagctt	1018
Db	12035	GCTTCGACTACATGGTTACTCTTTTAATATACAAAAAACCTTGGGAACAATGCTAAATTA	11976
QY	1019	agagaagccttagcatcagccttagataaaggagcctctgtctcaactatatcttaggc	1078
Db	11975	CGCAAGCATTTAGCCTTTGCNATAGACAAAGATATGTTACCAAAAGTGTTATACCAAGT	11916
QY	1079	cgtgaaaaactgccgatcatctcctaccacaataattcaatgatctccgaacatcaa	1138
Db	11915	CTTGCAGAACCTACAGATCATATCTTACATCCAAGACTTTATCCAGGAGCCTATCCGAA	11856
QY	1139	aaacagagatggcaacagccaaagctctacgctaaaaaacttttaagaagcttttagaa	1198
Db	11855	CGMAAGACAAACGAAAGAAATTTCTGAGGCTCCAACTCTTTGAAGAAGCTCTAGAC	11796
QY	1199	gaactccaaactcaactgctaaagatctccaaacatcttaactctatctctccggttctcog	1258
Db	11795	GAACTTCAAAATGACACGCGAAGATCTAGAAAGGAACCTTTGACTTTCTCAACCTTTTCT	11736
QY	1259	tcagaagttctttactagtcacaacttatcagagaacagtggaagaaagtcttagggttc	1318
Db	11735	TTTCTTTACGGAAGGATTTGCCAAATGCTAAGAGAACAAATGGAAGAAGTCTTAAATTT	11676
QY	1319	gcataccctattgctggaagaagtctgctctctccaagcagacctatctcagggaac	1378
Db	11675	ACTATCCTTATAGTAGGCCAAGATTTTTCAAATACAAAAAACTTCTCTAGAGGGGAAC	11616
QY	1379	tctcttttagctacagaggatggctgcagactttgctgacctatggcatttctaag	1438
Db	11615	TATTCCTTAACCGTGAACCAATGGACCGACCACTTTATTGATCCGATGCTTATCTCATG	11556
QY	1439	atcttgcttatcaatcagagttctctcttatgcaataccataaggacttcttagaa	1498
Db	11555	AUCTTTGCCAATCTCGAGGAATTTCCCTTATCACTCCGAAGATTACACATTTCAAACT	11496
QY	1499	attctacaacaatagaacaagcagaagatccacaaaacgctcggaattagtgcgcaa	1558
Db	11495	CTTCTATAAAGATCACTCAAGACATATAAAACACCTTACGAATACGCTTTATTTGAA	11436
QY	1559	gcttcttttaactagagaccttcaattattatgagcagatctaccacagcagctttcaa	1618
Db	11435	GCCTTTGACTATTAGAACACTGTCACTTCTCGAACCACTATGTCAATCCAAATCTTCGA	11376
QY	1619	ttagctatgaataaaaaactttctaactctagaggtctcaccacagaggtatggaactc	1678
Db	11375	ATTGCTTGAAACAAAACATTAAACATTTATCTCTTTTGTTCGACGACTTCAGACTTT	11316
QY	1679	cgttatgctaaggaaattagca	1701
Db	11315	CGTTTTATAGAAAAACTATAGGA	11293

RESULT		8
AX349605		
LOCUS	AX349605	1584 bp DNA
DEFINITION	Sequence 128 from Patent WO020606.	linear
ACCESSION	AX349605	
VERSION	AX349605.1	GI:18615409
KEYWORDS		
SOURCE		
ORGANISM		
	Chlamydophila pneumoniae.	
	Chlamydophila pneumoniae	
	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.	
REFERENCE	1 (sites)	
AUTHORS	Ratti,G. and Grandi,G.	
TITLE	Immunisation against Chlamydia pneumoniae	
JOURNAL	Patent: WO 020606-A 128 10-JAN-2002;	



FEATURES	Chiron S.p.A. (IT)	Location/Qualifiers
Source	1. .1584	
	/organism="Chlamydoiphila pneumoniae"	
	/db_xref="taxon:83558"	
BASE COUNT	506 a 378 c 256 g 444 t	
ORIGIN		
Query Match	22.8%; Score 409.4; DB 6; Length 1584;	
Best Local Similarity	54.8%; Pred. No. 3.2e-89;	
Matches 834; Conservative	0; Mismatches 681; Indels 6; Gaps	
QY 185	tctctacatctcggggagaactcgtattataataagagatgaaccccgctctttagat	244
DB		
64	TCCCAATCCTCTAAAGGAATTTTGTGGTAAATATGAGGAAATGCCACGCTCCTTGGAT	123
QY 245	ccaagacaagtgcgacctctctcgaatacagccttgcaacaatatctatagaggatta	304
DB		
124	CTGTGGAATAACTCGTCTCATTTGCAGACAAACTTAATGCGTCATCTATATGAAGGACTC	183
QY 305	gttcaagaanaatacttttcagagaaatagagcctgctcttgcaagaactactctt	364
DB		
184	GTCSGAAGAACAATCCCAAAATGGAGAGATTAAACACAGCCCTTGCAGAAAGCTACACATC	243
QY 365	tctctgcagcagcactcaactatacttttaaacactgaatacagctttttggagtaatggcag	424
DB		
244	TCCGNAACGCGGACTCGGTACACATTTTAAATCAAAAACATCTCTTTGGAGTAACGGAGAC	303
QY 425	cccttaacagctgaagactttatagaactcttggacaacagtagctactcaagaagctctca	484
DB		
304	CCTCTGACAGCTCAAGACTTTGTCTCTCTTGGAGGAAATCTCTAAGGAAGATGCGGTCC	363
QY 485	ggaaactatgcttttgcctctgaatccaattaaaaalgtacgaagaatccaagagggacac	544
DB		
364	TCCGTATATCTCTATGCGCTTTTACCTATCAAAAATGCTGGCAATCTTTGATGATACT	423
QY 545	ctctcccatagacattttggagtcgactctcctaataatgaattacactctgttgttaccctg	604
DB		
424	GAGTCTCCAGAAATCTAGGAGTCCGAGCTTTAGNATAGGCGTCATCTCGAATTCAGTTA	483
QY 605	gaatccccaactcgcattctttaaacttttagctctctccagcttttttcccggttcat	664
DB		
484	GAAACTCCCTGGCGCATTTCTCTACATTTCTTGACTCTTCTATTTTTCCTCCGTTCAAT	543
QY 665	aaatctcaagaagacctgcactccaactctota-----cctatagcagcgaggettctc	718
DB		
544	GAAACTCTCGAAGACTATAGCACTCTTTTGAAGAGATGCCCATTTACCTGGCGGTGTTTC	603
QY 719	tatcttaaaaatatcaacaaaaaacatgataaactctcaaaaaacctccactact	778
DB		
604	CGCCCTGTCTCTAGAAAAGCCCTGAGACTCCATCTAGAGNAAAACCCCTATGTACCAAT	663
QY 779	aatcaaatcaggttggaactaaaacgattacgatttoacttctcccgatgcaaacaca	838
DB		
664	AATAAAGCCGTGTAACACTACATAAAATATTCTACAGTTATCTCAACGCTTAACACT	723
QY 839	gcagcaaatatttaactcagggaaaactcgaattgcgaaggacctctctggggagagacgc	898
DB		
724	CGAGCCATTCTTCACACATATGAAGAAATAGATTGGCAGGAGACTCCTTGGGGAGAACCT	783
QY 899	attcctcaagaanaacctatccaatttacagtcctaaaggcgacttaacactcttttgatctc	958
DB		
784	ATCCCTCCAGAAATCTCAGCTTCTCTACATCAAGATGACCACCTCTTTCTCTCCGGCG	843
QY 959	gcaggaaactaatggtccaccttcaataatcaataattccccctcaacaatatgaagctt	1018
DB		
844	GCTTCGACTACATGGTTTACTCTTTTAAATACAAAAAAACCCTTGGCAACATGCTAAATTA	903
QY 1019	agagaagccttagcatcagccttacaataagaagactcttgctcaactatatattctaggc	1078
DB		
904	CGCAAGGCATTGACGCTTGCATATAGACAAAGATATGTTAACCAGTAAGTGTATACCAAGGT	963

Qy	1079	cgtgcaaaacgcccgcatactctcctacataattattctagctatcccgaaacataa	1138
Db	964	CTTGAGAACCTACAGATCATCTACATCCCAAGACTTTATCCAGGACCTATCCCGAA	1023
Qy	1139	aaacaagagatggcaacagccagcttaccgctaaaaacttcttaagaagcttttaga	1198
Db	1024	CGGAAAAAGACAAACGAAGAATCTTTGAGGCTCAACAACCTCTTTGAAGAAGCTCTAGAC	1083
Qy	1199	gaactccaaactcactgctaaagatctcgaactcttaattctattcttcccggttccctcg	1258
Db	1084	GAATCTCAATGACACGCGGAAGATCTGAGAAAGGAACTTTGACTTTCTCAACCTTTTCT	1143
Qy	1259	tcagcgaagtcttttactagtcocaaacttatacagacagcagtggaaagaagcttttagggttc	1318
Db	1144	TTTTTCTTCGGAAGGATTTGCCAAATGCTTAAGAGACAAATCGAAGAAGTCTTAAAAATTT	1203
Qy	1319	gcataccctattgtcgaagaagaattgctcttcccaagcagacctattcttcagggaac	1378
Db	1204	ACTATCCCTATAGTAGGCCAAGAGTTTTTACAAATACAAAAAACTTCTCTAGAGGGGAAC	1263
Qy	1379	ttctcttagctacaggaggatggttcgacagcttctgctatcctatgcatcttcgaattctcaag	1438
Db	1264	TATTCCTTAACCGTGAAACCAATGGACCGCAGCATTTATGATCCGATGCTTTATCTCATG	1323
Qy	1439	atcttggcttatccatcaggagttctctcttgcataatcaaccataagcacttctctagaa	1498
Db	1324	ATCTTTGGCAATCCTGGAGGAAATTCGCCCTATACCTCCCAAGATTCACACTTCCAACT	1383
Qy	1499	attctacaaaacatagaacagaagcagaatcaccacaaacgcctcggaatagtcgcaa	1558
Db	1384	CTTCTCATTAAGATCACTCAAGAACATAAAACACCACTCAAGAAATCAGCTTATTATTGAA	1443
Qy	1559	gctctcttcttaactagagaccttcttcatattatgagcgcgatctaccacgaagcatttcaa	1618
Db	1444	GCCCTTGACTATTAGACACTCTCAATCTCGAACCACTATGTCTCCAAATCTCGA	1503
Qy	1619	tttgctatgaataaaaaoattcttaactatagtaggtctcaccacagaggttgtagcttc	1678
Db	1504	ATGTCTTTGAACAAAAACATTAATAACTTTAATCTTTTGTTCGACGACITCAGACTTT	1563
Qy	1679	cgttatgctaaaggaataatag	1699
Db	1564	CGTTTATAGAAAAACATATAG	1584
RESULT	9		
LOCUS	AE002315	12173 bp	DNA linear BCT 26-MAY-2000
DEFINITION	Chlamydia muridarum, section 46 of 85 of the complete genome.		
ACCESSION	AE002315		
VERSION	AE002315.2		
KEYWORDS			
SOURCE	Chlamydia muridarum.		
ORGANISM	Chlamydia muridarum		
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.		
AUTHORS	1 (bases 1 to 12173) Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.		
TITLE	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39		
JOURNAL	Nucleic Acids Res. 28 (6), 1397-1406 (2000)		
MEDLINE	20150255		
PUBMED	10684935		
REFERENCE	2 (bases 1 to 12173) Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.		
AUTHORS	Direct Submission		
TITLE			



JOURNAL Submitted (01-MAR-2000) The Institute for Genomic Research, 9710  
 COMMENT Medical Center Dr. Rockville, MD 20850, USA  
 FEATURES On Jun 1, 2000 this sequence version replaced gi:7190506.  
 source Location/Qualifiers  
 1. 12173  
 /organism="Chlamydia muridarum"  
 /db\_xref="taxon:83560"  
 /note="synonym: Chlamydia trachomatis MoPn"  
 complement(97. .792)  
 /gene="TC0454"  
 complement(97. .792)  
 /gene="TC0464"  
 /note="hypothetical protein; identified by Glimmer2;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAF39314.1"  
 /db\_xref="GI:7190507"  
 /translation="WDSCALKNLSCEVSYRIPVIATRKREVALVILVIGTVLAAIG  
 ACVAVISCWMSIGLTVYGGVLLATGFLFALMEICRAQYGRQQRIOIQGVLESS  
 PIREATLPFIQAVYSHEPELPLSSQEEVRCRPIILQRGKPAETPKFIAVGSVV  
 ELVGMIGPMGRGNVNPQOTLVRWLDELALGRMELVRLDGFCECKVLPATLQDTS  
 VKTSADRASDED"  
 complement(1227. .2345)  
 /gene="TC0465"  
 complement(1227. .2345)  
 /gene="TC0465"  
 /note="similar to GB:AL009126; identified by sequence  
 similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="guanine tRNA-ribosyltransferase"  
 /protein\_id="AAF39315.1"  
 /db\_xref="GI:7190508"  
 /translation="MALRFEILHQSKSRARVRIETHEGICDTPAFVPVATNGALG  
 VLDHNTPLMFCNTYHLIVHGAZAVAMGGLHOFICRNPAPILITDSGGQIFSLAIGS  
 VAEIKSGKKKGNSIIKINDEGVFKSYGRGKRLSPSEVQACKLHGLADIIPL  
 DELLPFHPDPAITQSSORTYANERKSLDYHLANGKLSYGVHGGTFPDQRKLGCC  
 FVEDLPDGAIGSGSKNLRIOIVGVDTTANLSIERPHLLIGLDLPSIWAIVGFG  
 IDSFDSPYPTKARHGMLITSQSLKINRYASDLNPIEPCACPCACSQGISRAYLR  
 HLFVHPNAGIWSIHNNHYMQKIMSKIREKILNDQL"  
 complement(2467. .3939)  
 /gene="TC0466"  
 complement(2467. .3939)  
 /gene="TC0466"  
 /note="magnesium transporter; identified by match to TIGR  
 protein family HMM TIGR00400"  
 /codon\_start=1  
 /transl\_table=11  
 /product="magnesium transporter"  
 /protein\_id="AAF73559.1"  
 /db\_xref="GI:8163227"  
 /translation="MGEONRNEEKLVALSGNLSKTSKSHLDLSFKLEAFDTILT  
 AGBSODLTSIVSANPIDLAYVSCLPDSRSLYSLSDSIASKIAPFINIDSASRW  
 ALPNLSDEEICALIEQMPPEADLWLLDIPDRRYRRLDLIDVKKALKRLQKGR  
 NAGRKMTTEFAFLETMYTKEVATCIRNPNNGVDLRLVFLDKFGLQGVFDRSLI  
 IASPENLQIIMRPERVHKVLAADTTRVEVDLVKRYKAVLPVDEENFLGATYEDV  
 VETIEDIETARMAGTTEDGVHDCVQVQRELLRAPWLLITLCAGLVSASVMAVYFQ  
 KIAPTLAMVIFPIFVNLGSLNMGVQCSTILVRSMTATGTSFGRRRETILKEMSGL  
 LTVALGILGVLCMCGCLGLFTGGVQVGVTVSVGILGASLTATLGLVLSPPFF  
 AKIGVDPALASGPVITALNDINSMVIFLTITGLNVLFFK"  
 4208. .4366  
 /gene="TC0467"  
 4208. .4366  
 /gene="TC0467"  
 /note="hypothetical protein; identified by Glimmer2;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAF39316.1"  
 /db\_xref="GI:7190509"  
 /translation="MKAHTVHPFPOLQLYDEDFTVDLQAKSLGLLKTAVVYAKLAAS  
 QPSQSGPI"  
 complement(4328. .5410)  
 /gene="TC0468"  
 complement(4328. .5410)  
 /gene="TC0468"  
 /note="conserved hypothetical protein; identified by  
 Glimmer2; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="conserved hypothetical protein"  
 /protein\_id="AAF39317.1"  
 /db\_xref="GI:7190510"  
 /translation="MFLNLPFAEVLRPVTTSEFPRASSWKATQKRITSCITLTNPFYK  
 TSLTRIVCFVVGILIVLGAICRQIAYLANAGLLFSFKILAIALLSSVLLGGIGLCK  
 IARKIDVIGEKIQFAIRKWEQVITCEKRGQTIPIODPDMVMDSVLLDKQSGSIAP  
 VYTPPAKARTAICTACILLPITVTVLYNAFRELLIPFIVQIMRQLYQEDLPF  
 EEOPTGSDIPREMSRFSQAVKAPFYGVACYLASLYLLNPLUGSRVIMASVERDWNND  
 VTRSGWIGIFCEKNCLEGGTSGGLGQAHWYLLGCFQFRLFLLLKDGDEIVSGARPS  
 IQAPPENKELTISYLYGAALGRLTSR"  
 5581. .5946  
 /gene="TC0469"  
 5581. .5946  
 /gene="TC0469"  
 /note="hypothetical protein; identified by Glimmer2;  
 putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAF39318.1"  
 /db\_xref="GI:7190511"  
 /translation="MNYLFSLNQKLRKMQLTIVNTNSRIFFSPSNERILPGSTLSTD  
 STFSYKILIPLLAGSICLLGALLAGTITTPAIAASYELSLGVLVVGIGLCSV  
 FRPHLFSINQSDLIHYIY"  
 6012. .7034  
 /gene="TC0470"  
 6012. .7034  
 /gene="TC0470"  
 /note="O-sialoglycoprotein endopeptidase; identified by  
 match to PFAM protein family HMM PF00814"  
 /codon\_start=1  
 /transl\_table=11  
 /product="O-sialoglycoprotein endopeptidase"  
 /protein\_id="AAF73560.1"  
 /db\_xref="GI:8163228"  
 /translation="MUTLGLSESCDETSCALVENGKILANRIASQDIHAAYGVIPEL  
 ASRAHLQFLPKLAADAEVSELDVELISVANTPGILGALSVGVNFAKGLASGLKK  
 TILGVNHYEAHLAAACLEPSPIRFPALGLIAGHTSLIFLMPNATTFLLICKTRDDAI  
 GETDKVARFLGLPYPGOKLEELAQDGDDEAYPFSRAKYSNGDNFSGLKATVLIYAL  
 KGNSSAKAPPEVSETQKNTIAASFQKAAFWTIAQLPDIVKAFSCSLIVGGGVAN  
 NRYPRLLNQCSLPTPYFFSSQLCSDNAMTAGLGEKLFQNTYVSKEVIPCARYQWE  
 SACLSH"  
 6992. .8554  
 /gene="TC0471"  
 6992. .8554  
 /gene="TC0471"  
 /note="similar to GB:M57689 SP:P24141 GB:X56347 PID:143603  
 PID:40005; identified by sequence similarity; putative"  
 /codon\_start=1  
 /transl\_table=11  
 /product="peptide ABC transporter, periplasmic  
 peptide-binding protein, putative"  
 /protein\_id="AAF39319.1"  
 /db\_xref="GI:7190512"  
 /translation="MRKVSVGIOLLIATAITCCKSSKSNHSSNOSVSVSKMD  
 DPRTQPREVRLLSIDNLIHHLYEGVLQVETPSGEVFPALAESFLLSEKTYTFHLKK  
 ALMSNGDLITAHDFVRSWNVLONRIASIVSFAFLPIDLKDCCGFFAKDNHTLVINLH  
 TPTPHEKLLTLTPVFPVPHPEQIRNEAKALPISTGAFCLCKDRKWLKKNPYYY  
 NKEQVAIOEIHIVIPDOQTASALFKQGLDWOGUPWHSIPQALATANKRTPQSF  
 DISGTQWLFNTSKPFFSHPKRLQALSLKLEALASPTFYKPAKHLIPLTHLTYPEQ  
 PTYKQAEVILLATLLQALDNLMTIKOLEKPLIFTSATSSVNSQAOAMRDQWERI  
 LGTIPFCGKEYALLNDLNTTFMFSINGNADFSDPLAFSLVFSKGIKPYALQDP  
 LFTQLILSTETEKDPKRKISLISBASLYLEKQNIIEPLYHDVYFAANKLSFVRLHP

gene  
SGLVDMRYAKNS"  
8668...9612  
/gene="TC0472"  
CDS  
8668...9612  
/gene="TC0472"  
/note="similar to SP:P24138 GB:X56347 PID:580897  
GB:AL009126; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="peptide ABC transporter, permease protein"  
/protein\_id="AAF39320.1"  
/db\_xref="GI:7190513"  
/translation="MLGYIKKLLFNLLSVVVLFFIIRIPGDPNDENGVL  
BELAILNKRGDLEFFQYLIYLLKLLTLDGESLIYKDRIVIGIITLPSAIL  
GLESILALFGGTLIGLAAFKRGCGRTIFFSIIQISVPAPVIGAFLOYIFAIKYS  
LLPTACWGDPSHLLPSIALAPMAFITQIYVSASLKKDYLLAVAGLPLKY  
LIRKVLPAFLFVIVSYSAFLITLMTGTFSENLCFIPGLGKWFICSKQRDYPMTLG  
LSVFGYAFFMLISLLCDLLQAWIDPQIRYSVGRHSKQ"  
9618...10481  
/gene="TC0473"  
9618...10481  
/gene="TC0473"  
/note="similar to GB:M57689 SP:P24139 GB:X56347 PID:143606  
PID:40007; identified by sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="peptide ABC transporter, permease protein"  
/protein\_id="AAF39321.1"  
Query Match 21.9%; Score 393.8; DB 1; Length 12173;  
Best Local Similarity 55.4%; Pred. No. 1.7e-85;  
Matches 942; Conservative 0; Mismatches 702; Indels 57; Gaps 7;  
QY 1 actccccctgctaactatgctagataatgctgctattgattgaggtcttagggggag 60  
Db 6889 ATTTCCTCTTCGCACTTGTCTCTGATATGCTGCAATGATAGCTGGATTAGGAGA 6948  
QY 61 aaaaatttcaaaaaactctagattcc---ggaattcgtatgataagagatcagat 117  
Db 6949 GACTGTTTGCACATACACATATGTTCTTAAGAGAGGTATTTCATGCGCAAGGTATCACT 7008  
QY 118 ggaatctg-----tatcaccattctcttagcctctccgttagctctcgaaggctgaa 171  
Db 7009 GCGAATCTGCTTGTGTTATGCTATGACACACACAAATACCGGCTGCTCAAAATCTCTC 7068  
QY 172 ggaagcagtcactctctacatctcggggggaactcgtattataataagaagatgaacc 231  
Db 7069 AAACAAATCAAAATCACTCTCTCTCTAATCAAAAGCGTATCTGTGAGCATGAAGACGATCC 7128  
QY 232 ccgttctttagatccaaagacaaagtcgactctcttcagaatacagccttgctcaaacatat 291  
Db 7129 TCGCAGGTTTGACCTCGAGAGGTAGGCTTCTTCGGATATTAATTTGATTCAACCATCT 7188  
QY 292 ctatgaggattagtttaagaataataattcttcaggaaatatagaccctgctcttgaga 351  
Db 7189 TTATGAGGAGCTGGTTCAGG---AGACTCTCTCCGGAGAGGTTTCCCTGCGATTAGCAGA 7245  
QY 352 agactactctcttcctcgagaggaactcaattatactttttaaactgaaatcagcttttg 411  
Db 7246 AGHTTCTCTCTATCTGAAGATAGAAACACCTACACTTTCACCTAAAAAAGCACCTG 7305  
QY 412 gagtaaggcagcccttaacagctgaagactttatagaactttggaacaagtagctac 471  
Db 7306 GAGTAATGGAGACCTGATTTACTTGCTCATGATTTTGACGTTCTGGAAGCATGTTTGA 7365  
QY 472 taagaagctcaggaaatctatgctttgcttgctgaatccaataaaatgaacgaagat 531  
Db 7366 AAATCGCATTCGAATGATCTACTCTTTCGCTTTTATACCATGTAT----- 7411  
QY 532 ccaagaggacacctctccatagaccatttttgagtgagtcactctctaatgaatcact 591  
Db 7412 -----TTAAGTAAAGATTTGTGGATTCTTCGTAAGGACATATACACT 7455

QY 592 tgtgttaccctggaatcccaactcgcattttcttaaaacttttagctcttcaggtctt 651  
Db 7456 TGTTATCAATTTGCATACCCCTACTCCACACTTTCTTGAAATTTACCTTCCAGTATT 7515  
QY 652 ttccccggttcata---aatctcaagaacccctgcaatcccaaatctctacctatagaag 708  
Db 7516 TTATCTGTGTCACCGCGACATCAATTCGGAATGAGCAAAAGCTCTTCCAATATCTAC 7575  
QY 709 cggagctttcttacttaataataatcaaaacaaatggataaaaactctcaaaaacc 768  
Db 7576 AGGGGCTTTTGTGTTTAAAGAGAAAAAGATCGCGCTTAAATTTAGAAAAATCC 7635  
QY 769 tcaactataatacaagtcagtggaactaaacgaatcaacgaatcaactcattcccgga 828  
Db 7636 CTATTTATTAACAAGACAGAGTTTCCGATCCAGAAATCCACATACATGTGATTCTGA 7695  
QY 829 tgcacacacagcagcgaactttaaactcagggaactcaattggcgaaggaactccttg 888  
Db 7696 TCAGCAGACCGCTTCTGCTTTTGTTCACAAAGGAAATTTGGATTGGCAAGGACTTCCTTG 7755  
QY 889 gggagaacgattctcctaagaacccctatccaatttaacagtcctaaggggcacttacact 948  
Db 7756 GGGACACAGCATTTCCCAAGAAAGCTTTAGCAACAGCAAAATAAACGTCGTACTCTCAATC 7815  
QY 949 ttttgatgctgcaggaacccctcaatggctcaactcaatataataaattcccccacaac 1008  
Db 7816 ATTTGATATATCTGGCACCCTGCTGGCTCAGTTCAACACATCAAAAGTACCTTCAGCCA 7875  
QY 1009 tatgaacttagagaagccttagcatcagccttagataaaggagctcttctgctcaactat 1068  
Db 7876 TCCCAAGTTGCGACAGCTTTAAGCCTTTGTTTAAATAAGGAAGCTTTAGCCTCTCAAC 7935  
QY 1069 attcttagccgtgcaaaactcgcgatactctcctacatacaataatctcatagctatcc 1128  
Db 7936 CTT-----TGTAGCCGCGCTTAACACCTACTTCTACACACCTGACACCTTATCC 7986  
QY 1129 cgacatcaaaaacagagatggcacacgccagcttaagcttaaaactctttaaaga 1188  
Db 7987 AGAGCAACCTACTTACAACACAGAGCGCTCATTTTAGCCAAAATCTGTGTACAAAGA 8046  
QY 1189 agotttagaagaactccaaatcactgctaaagatctcgacatcttaactctatcttcc 1248  
Db 8047 AGCATTAACAGATCTTAATATGACTATAAAGATCTTGAATAATGCCCTTTGATTTTTC 8106  
QY 1249 cgttctcctgcgcagcaagctcttacttagtccaaactatacagagaacagtggaagaag 1308  
Db 8107 AGGCACATCTACGTCGCACTCACAATATGCCCCAAATGATGCGCGATCAATGGCGAAGAT 8166  
QY 1309 tttagggttcgctatccctattgctcgaaaggaatttgctcttctccaagcagaacctatc 1368  
Db 8167 TTTAGGAATAACATTCCTATTTTCGGGAAAGAGATATGCTTTTATACAAATGATTTGAC 8226  
QY 1369 ttcagggaactctctttagctacagaggatgggttcgcagacttttgctgactctatgsc 1428  
Db 8227 CACCAACACTTTTTTATGTCCATAAACGCTTGGTTTCAGACTTTTTCOGATCCTTTAGC 8286  
QY 1429 attcttaacgacttttcttattccatcagaggatctcctccttatgcaatcaacaaagga 1488  
Db 8287 CTCTCTGCTGTTTTTTCTT---CTAAGGGAATCAAACTTATGCTTTTACAAAGACCTCT 8343  
QY 1489 ctcttagaattctcaaaaacatagaacagacagatcaccacaaacgctcgaatt 1548  
Db 8344 GTTTCACGAGCTTATCTCTATAGAAACGGAAGAAAGATCTCGAAAGCGTATTTCTCT 8403  
QY 1549 agtgcgaagctctcttcttaactagagacctttctattattagccgatctacacga 1608  
Db 8404 GATTTCGCAAGCTTCTATACCTAGAAAAACAAAATATCATAGAACCCCTCTATCAGA 8463  
QY 1609 cgcatttcaatttgctatgaataaaaaactttcttaacttaggaggtctcaccacaggagt 1668  
Db 8464 TGTTTTTCACATTCGCGCAAAATAAATAAACTTTCTTTTGTAGATTACATCCCTCAGGCT 8523  
QY 1669 tgtggacttcgcttatgctaa 1689

```

Db 8524 TGTGGATATGCGGTATGCAAA 8544
|||||
RESULT 10
AX300393
LOCUS AX300393
DEFINITION Sequence 5 from Patent WO0185972.
ACCESSION AX300393
VERSION AX300393.1 GI:17381784
KEYWORDS
SOURCE
ORGANISM Chlamydomophila pneumoniae.
REFERENCE 1 (sites)
AUTHORS Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
TITLE Chlamydia antigens and corresponding dna fragments and uses thereof
JOURNAL Patent: WO 0185972-A 5 15-NOV-2001;
Aventis Pasteur Limited (CA)
FEATURES
Location/Qualifiers
Source 1..1235
/organism="Chlamydomophila pneumoniae"
/db_xref="taxon:83558"
101..1135
/feature="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAD13019.1"
/db_xref="GI:17381785"
/transl_table="MULTIGLSSCDTACATVNDKQILANLIASQDTHASXGVVPE
LASRAHLHIEPQVINKALQOANLIEDMDLIANTQPPGLISLVGVHFGKGIAGAK
KSLGVNVHVEAHLYAYMAAONVQFPAIGLVVGAHTAAFFIENPTYSKLGKTRDDA
IGETFDKGVGFLPYAPAGPLIEKLEGGSEDSYFPSPAKVFNDFSFSLKTLVLA
IKGNSSPSRPAEISLEKQDIAAFKAACTTIAOKLPTIKFESRSLIGGVA
INEFRAIQTAQCNUPVFPVPAKLCSDNAMIAGLGFENFKNSSIPRIRICARYQE
SVSPFSLASP"
BASE COUNT 365 a 278 c 230 g 362 t
ORIGIN

Query Match 14.2%; Score 255; DB 6; Length 1235;
Best Local Similarity 100.0%; Pred. No. 1.4e-51;
Matches 255; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 acttccccctgctaaactatgctcagataatgctgctatgctgaggtcaggtcaggtgag 60
|||||
Db 981 ACTTCCCCCTGCTAACTATGCTCAGATATGCTGCTATGATGCTAGTTCAGGCTAGGCGGAG 1040
|||||

Qy 61 aaaaatttcnaaaaaacttagtattccgggaattcgatatgctgcaagaatatacagtgag 120
|||||
Db 1041 AAAAATTTTCAAAAAAACTCTAGTATTTCGGGAAATTCGTATATGCGCAGATATCAGTGGG 1100
|||||

Qy 121 aatctgtatcaccattctcttagctctccgtagctcccaaggctgcaaggagtcagag 180
|||||
Db 1101 ATATCTGTATACCAATTCCTCTAGCCTTCGCTAGCTCTCCTCAAGGCTGCAAGGAGTCCAG 1160
|||||

Qy 181 tcaactctctacatctcggggagaactcgtctattaataagaagatgaaccccgctctttt 240
|||||
Db 1161 TCACCTCTCTACATCTCGGGAGACTGCTATTAATATAAGAGATGAACCCCGCTCTTT 1220
|||||

Qy 241 agatccaagacaagt 255
|||||
Db 1221 AGATCCAAGACAAGT 1235
|||||

RESULT 11
AX349505
LOCUS AX349505
DEFINITION Sequence 28 from Patent WO0202606.
ACCESSION AX349505
VERSION AX349505.1 GI:18615359
KEYWORDS
SOURCE Chlamydomophila pneumoniae.

```

```

ORGANISM Chlamydomophila pneumoniae
REFERENCE 1 (sites)
AUTHORS Ratti,G. and Grandi,G.
TITLE Immunisation against Chlamydia pneumoniae
JOURNAL Patent: WO 0202606-A 28 10-JAN-2002;
Chiron S.p.A. (IT)
FEATURES
Location/Qualifiers
source 1..1587
/organism="Chlamydomophila pneumoniae"
/db_xref="taxon:83558"
BASE COUNT 547 a 397 c 238 g 405 t
ORIGIN

Query Match 10.1%; Score 181.4; DB 6; Length 1587;
Best Local Similarity 49.3%; Pred. No. 1.1e-33;
Matches 640; Conservative 0; Mismatches 626; Indels 33; Gaps 5;

Qy 202 agaactcgctattataataagagatgaaccccgcttcttagatccaagaagtcgagct 261
|||||
Db 111 ACATCTCGTTATTTCGGATGAGCATGATCTGCGGACCTAGATCTCTCGCAATGCTATT 170
|||||

Qy 262 tctttcagaatacagccttgcataacatatctatgaggaattagttccaagaaataatct 321
|||||
Db 171 AAGCAGAGATGCTTCCTCTAGCAAAAGCCCTCTATGAAGGACTGACAAGAGAACTGATC- 229
|||||

Qy 322 ttcgggaataatagacgtcgtcttcgagaagactactcttcttcgagcgagctcac 381
|||||
Db 230 -----AAGGAATCGCACTGGCTTTCGAGAAGTTATACCTCTCAAAAGATCATAGGT 284
|||||

Qy 382 ttactctttaactgaatacagctcttttgagtaataaggcacccttaaacagctgaaga 441
|||||
Db 285 CTATACCTTTAACTCAGACCTTCTGTGTGGAGCGATGCGACCTCCACCTCACTGCTTATGA 344
|||||

Qy 442 ctttatagaatcttggaaacaaagtagctactccaagaagctctcaggaatctatgctttgc 501
|||||
Db 345 CTTTGAATAAATCTATAAAACAACGTACTTCGAAGAATTTTCACCTTCATACATACTTT 404
|||||

Qy 502 ctggaatccaataataaataatgtacaaagatccaaggaggaacacctctccatagaccattt 561
|||||
Db 405 ACTCGGCGTGAATAAAATTTTCGGCAATCCCAATGCTCAAAATCTCTGGAACCTCT 464
|||||

Qy 562 tggagtgacactctcctaataatgaatcactgtgtgttacacctgggaatcccaacctcgca 621
|||||
Db 465 TGGGATACAGGCAAAAGATGATCTTACTTTGGTGATACCCCTAGAGCAACCTTTCCCATATA 524
|||||

Qy 622 ttctttaaacttttagctcttccagctcttttcccgcttcata-----aaac 669
|||||
Db 525 CTTTCTCACACTTATCGCTCGCCCGCTATTCTCCCGCTGTTTCATCACACCCCTTAGGGAATC 584
|||||

Qy 670 tcaagaacacctgcaatccaataatcttaacctatagcagcgaggtcttctatccataaaa 729
|||||
Db 585 CTATAGAAGGAACACCCCGCTACATACATCTCCCAATGCGGCTTTGCTTTAAAAA 644
|||||

Qy 730 tatcaacaaaaaacatgataaaactctcaaaaaacctcaactactaatatacaagatca 789
|||||
Db 645 ACATGACACCAAAACTACTTAATTTAGAAAAAATCTCTCACTACTATGATCATGAATC 704
|||||

Qy 790 ggtggaaactaaaacgattacgattcacttctcccgatccaacacacagcagcaaaact 849
|||||
Db 705 AGTAAAGTTAGACGAGTACCTTTAAATTTATCCAGACGCCCTCCACAGCCAGCAAACT 764
|||||

Qy 850 atttaatacagggaaaaactcaattggcaaggacctctcttggggaacgacttctcaaga 909
|||||
Db 765 TTTCAAAAGTAAATCTATAGATTGGATTGGCTCACTTTGGAGCGCTCCGATATCTTAAGA 824
|||||

Qy 910 aacctatccaatttaacgtcttaaggggcaacttaacactcttttgatgctcgaggaactc 969
|||||
Db 825 AGACCAA---AAAGTTCTCTCCCAAGAAAGATTCTTACCTATTCTGTTTCAAGCAACC 881
|||||

Qy 970 atggctcactcttaataatcaataaattccctcccaaatatgaacttagagaagcctt 1029
|||||

```

Db 882 CTTCTTATATACCTGCAAAACCTCTAATACAAATTAAGCCCTCAGCAAGCCAT 941  
 QY 1030 agcatcagcttagataaagaagctcttgctcactatattcttaggcgtgcgaataac 1089  
 Db 942 TGCTCATGCTATTGATGAGAAATCTATCTTAAGACTCGTGCCTTCAGGACAGAA----- 996  
 QY 1090 tgcgcacatctcctacacataattcatagctatccgcgaacatcaaaacaagagat 1149  
 Db 997 -GCTGTAACCTTAGTTCCTCCCAATCTTTCACAACTCAATCTTCAAAAGAGATCTCAAC 1055  
 QY 1150 ggcacaacgcagcttagctgtaaaacactctttaagaagcttttagaagaactccaaat 1209  
 Db 1056 AGAAGAACGACAAACAAAGCCAGAGCATATTTTCAAGAGCTAAGAAACAC-----T 1109  
 QY 1210 cactgcctaaagctcgcgaacacttaattcttcttccgcttctcctcgcgaagcttc 1269  
 Db 1110 TTCTGAAGAAGAACTCGCAGACTCAGCATCTCTATCTATAGATTTCTCGAATTCCTC 1169  
 QY 1270 ttactagtcacacttaccagacagctggaagaagcttttaggcttgcgtatccctat 1329  
 Db 1170 CATCATAGCTCAAGAAATCCAAAGACAACTTAAGATACCTTAGGATGAAATCAAAAT 1229  
 QY 1330 tgcggaaggaattgctctctccagcagacacttatcttgaaggaactctcttttgc 1389  
 Db 1230 CCAAGCAGTGGAGTACCACTGCTTTTAAAGAAACGCTCGCAAGGAGATTTCTCATAGC 1289  
 QY 1390 tacagagagatggttcgcagacttgcgtatgctctatggaacttcttaacagatctttgtta 1449  
 Db 1290 GACAGAGGATGGATTGGGAATACGTAAGCCCGTAGCCTTCTCTATCTATCTTAGGCAA 1349  
 QY 1450 tccatcagaggtctcctcttgaatcaatcaaccataagga 1488  
 Db 1350 CCCAGAGACCTCACACAATGGAGAACAGTGAATACGA 1388

RESULT 12  
 AX300389  
 LOCUS  
 DEFINITION Sequence 1 from Patent WO0185972.  
 ACCESSION AX300389  
 VERSION AX300389.1 GI:17381780  
 KEYWORDS  
 SOURCE Chlamydomophila pneumoniae.  
 ORGANISM Chlamydomophila pneumoniae  
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.  
 REFERENCE  
 1 (sites)  
 Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.  
 Chlamydia antigens and corresponding dna fragments and uses thereof  
 Patent: WO 0185972-A 1 15-NOV-2001;  
 Aventis Pasteur Limited (CA)  
 FEATURES  
 Location/Qualifiers  
 1..1787  
 /organism="Chlamydomophila pneumoniae"  
 /db\_xref="taxon:83558"  
 101..1687  
 /note="unnamed protein product"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAD13017.1"  
 /db\_xref="GI:17381781"  
 /translation="MKMRLKPTLKSLIPNLLFLLLTSSCKQKQPLGKHVIVAMS  
 HDLADIPNAYLSRDSALAKALVEGLTRETDOGTALALAESYTLKSDHKVYIFKLKP  
 SVNSGDPTLAYDFEKSIKOLYFEESPSIHLILGVINKNSAIHNAOKSLETGLIOAK  
 DDLTVITLQEPYFYLTLARVFSVPHVLTRESYKKGTPPTYSINGFPVLKKEH  
 QNVLLERNPHYDIESVKLDVTLKILIPDASTATKFKSIDWIGPSWAPISNEH  
 QVLSQELIYSVSTLLIYNLQRLKALKAIAHIDRKSILRLVPSGPEAV  
 TLVFNLSQMLQREISERTQKARAYFOEAKETLSKELAEILILYDSSNSII  
 ADEIQRDLKLYGLIKIOGMEVHCFLKRRQRODFIATGWTAEVSPVAFILIGN  
 PDELQWNSDYETLKLPHAYKENLKRAMEIIEETPIIPLYHGVIYIATHPKI  
 QNTFSLIGHTDLKNIDLS"  
 628 a 436 c 261 g 462 t

BASE COUNT  
 ORIGIN

Query Match 10.1%; Score 181.4; DB 6; Length 1787;  
 Best Local Similarity 49.3%; Pred. No. 1.1e-33;  
 Matches 640; Conservative 0; Mismatches 626; Indels 33; Gaps 5;

QY 202 agaatcgcctatttaataagagatgaaccccggtttcttttagatccaagacaagtgcgact 261  
 Db 211 ACATCTCGTTATTTCGATGAGCCATGATCTCGCGACCTAGATCCTCGAATGCCTATT 270  
 QY 262 tctttcagaatcagctctgcaacatatctatgagggattagttcagaagaataatct 321  
 Db 271 AAGCAGAGATCTTCCCTAGCAAAAGCCCTCTATGAAGACTGACAAGAGAACTGATC- 329  
 QY 322 ttcaggaataatagagcctgctcttgcaagaactactctcttctcgcgcgagactcac 381  
 Db 330 -----AAGGAATCGCACTGGCTCTTGCAAGAAGTTATACCTCTGCAAAAGATCATPAAGGT 384  
 QY 382 ttatacttttaaaactgaatcagctttttggagtaatggcgaccccttaacagctgaaga 441  
 Db 385 CTATACCTTTAAACTCAGACCTTCTGTGTGGAGCGATGGCACTCCACTCAGTCTGCTATGA 444  
 QY 442 ctitatagaatcttggaacaagtagctactcaagaagctctcaggaatctatgctttgc 501  
 Db 445 CTTTGAAGAAATCTATAAAACAACTGTACTTCGAAGAAATTTTCACCTTCCATACATAC 504  
 QY 502 ctggaatcccaattaaaaatgtacgaagatccgaagaggagacacctctccatagaccatt 561  
 Db 505 ACTCGCGTGTATTAANAATCTTCGGCAATCCCAATGCTCAAAATCTCTGGAATCTCT 564  
 QY 562 tggagtgacactctcctaatgaatctacactgtgtttaccctgggaatcccaacctcgca 621  
 Db 565 TGGGATACAGGCAAAAGATGATCTTACTTTGGTGATTAACCTAGAGCAACCTTCCCAT 624  
 QY 622 ttcttaaaacttttagctctccagctctttttcccggttata-----aatc 669  
 Db 625 CTTTCTCACACTTATGCTGCGCCGCTATTCTCCCTGTTCATCACACCTTAGGGAATC 684  
 QY 670 tcaagaacctctgcaatccaaatctctacctatagcaagcgagcttctctactctctct 729  
 Db 685 CTATAAGAAAGGACACCCCACTCCCATACATCTCCCAATGGGCCCTTTGTCTTAAAAA 744  
 QY 730 tatcaacaacaaatggatataaaactctcaaaaaacctactactataatcaagatca 789  
 Db 745 ACATGAACACCAAAACTACTTAATTTTAGAAAAAATACTCTACTACTATGATCATGAATC 804  
 QY 790 ggtlgaactaaacagattagactcactctcccgatgcaaacacagcagcaaaact 849  
 Db 805 AGTAAAGTTAGACCGAGTCACTTTAAAAAATTTATCCAGACGCTCCACAGCCAGAACT 864  
 QY 850 attaatcagggaacactcaattggcaaggacacctctctgggagaagcattcctcaaga 909  
 Db 865 TTTCAAAAGTAATCTATAGATTGGATTGCTCACCTTGGAGCGCTCGGATATCTAACGA 924  
 QY 910 aacctatccaatatttaacagtcaggggcactttacactcttttgatgctcgaggaacctc 969  
 Db 925 AGACCAA---AAAGTTCTCTCCCAAGAAAAAGATTCTTACCTATTCTGTTTCAGGCCAC 981  
 QY 970 atggctcaccttcaatatcaataaaattcccccctcaacaatatgaagcttagagaacct 1029  
 Db 982 CTTCTTATCTATACCTTGCAAAAGCCCTCTAATACAAATAAAGCCCTCAGGAAGGCCAT 1041  
 QY 1030 agcatcagcttagataaagaagctcttctcactatattcttagcgcgtgcgaataac 1089  
 Db 1042 TGCTCATGCTATTGATGAAATCTATCTTAAGACTCGTGCCTTCAGGACAGAA----- 1096  
 QY 1090 tgcgcacatcctcactacataatattcagctatccgaacatcaaaacaagaat 1149  
 Db 1097 -GCTGTAACCTTAGTTCCTCCCAATCTTTCACAACTCAATCTTCAAAAGAGATCTCAAC 1155  
 QY 1150 ggcacaacgcagcttagctgtaaaacactctttaagaagcttttagaagaactccaaat 1209  
 Db 1156 AGAAGAACGACAAACAAAGCCAGAGCATATTTTCAAGAGCTAAGAAACAC-----T 1209

```

QY 1210 cactgctaagatctcgaacatctattatctattcttcccgcttccctccgcaagttc 1269
DQ 1210 TTTCTGAAAAGAACTCCGAGAAGTCCAGCACTCTATCTATCTATAGATCTCTCAATCTCCTC 1269
QY 1270 ttctactagtcacacttatcagagacagcagtggaagaagtttgggttgcctatccctat 1329
DQ 1270 GATCATAGCTCAAGAAATCCAAAGACAACCTTTAAAGATACCTTAGGATTGAAATCAAAT 1329
QY 1330 tgcgtgaagaagaatttgcctctccacagcagaccctatctcagggaactctctctttagc 1389
DQ 1330 CCAAGGCATGAGTACCACTGCTTTTAAAGAAACGTCGCAAGGAGATTTCTTCATAGC 1389
QY 1390 tacaggagatggttcgacgaacttgcgtatcctatgcgaacttcttaacgatcttctgcta 1449
DQ 1390 GACAGGAGGATGAGTATCGGGAATAGCTAAGCCCGTAGCCTTCTATCTATTCIAGGCAA 1449
QY 1450 tccatcaggaggtctcctcttgccttgcctatgcgaacaccataaga 1488
DQ 1450 CCCAGAGAGCTCACACATGAGAAACAGATGATTAGCA 1488

RESULT 13
LOCUS AE002313
DEFINITION Chlamydia muridarum, section 44 of 85 of the complete genome.
ACCESSION AE002313
VERSION AE002313.1
KEYWORDS
SOURCE
ORGANISM Chlamydia muridarum.
REFERENCE 1 (bases 1 to 11402)
AUTHORS Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
TITLE Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39
JOURNAL Nucleic Acids Res. 28 (6), 1397-1406 (2000)
MEDLINE 2010255
PUBMED 10684935
REFERENCE 2 (bases 1 to 11402)
AUTHORS Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwin,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
FEATURES
    Location/Qualifiers
        1..11402
            /organism="Chlamydia muridarum"
            /db_xref="taxon:83560"
            /note="synonym: Chlamydia trachomatis MoPn"
        250..1569
            /gene="TC0440"
            /note="similar to GP:2444077; identified by sequence similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="phospholipase D family protein"
            /protein_id="AAF39294.1"
            /db_xref="GI:7190485"
            /translation="MCSPCPRLPPIVDCTPCSTPPLKRLPLSTTAAAVVWQLFK YHRKDLSHRVYDYGKNTKPRKIQCSNRRSLATSSKTSCKRILRRRTGSSVP RPIAQELFTISAQQTNSNPLKICKFDTRHQLHIQVCFNHPITIIQHLVKAAR REYRVSVQRDGTLEAEACKDSTICLQKSGSLFKKSWGDSORVLASGNFTPDS TEQDILSLINDSELASHITNCFVTRQIQFLVYPIPRNRKDFRLAVRTIKCI AQAKTIVVVVYILTHPVLKSLQAAAGVKVEAVDTRSEOTQTLRLQLSLPL

```

```

RVKPGAPRVHVMCYIDNKLICGSANWSSAGLTNRNREDFVIRGLTETQRQSLSEI
WQSVERTPLTAQSLKPREEDDPGEETSSGISAGASAKAKTQ"
complement(1661..3049)
/gene="TC0441"
/complement(1661..3049)
/gene="TC0441"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF39295.1"
/db_xref="GI:7190486"
/translation="MVDKFFFIQQLSLTSELSDPTDQFIKGLPKADIHVHLPLTSPETAWELGVNRGLIKMNQWSPAFSNGNPKHVAETIFNFESIERYERDPNVS GKAYITRDFADFRIMATVQGRHPPGGIQNENDLWLVLQNLKOCIOODNIYIEV QNIRIAYVYPSLKPLEARLRLYDLFSRASOLFISOGITRFLNCFNKTKSSNLOOS QTOORSEAAWLDDEAAPPNLVGLQSAGASCPEAPTKLTSGYHHAYENFGCEA HAGECTGFLYLNQTIQALPTORAHGFOAIEHVFTIHAIQENKTLVMAPAINLVGA SLHQYSGLEKISKTMNNLNDHPFFALFRDHKLVTLSNDPNMGSTSVQNTMLLSG FSAQDSHLTHVTSSPLTFKEITIQLANVEAIVSFVDVDTKIMLLNSIHEYLSYLIK KQYNLLEKKAFL"
complement(3064..4602)
/gene="TC0442"
/complement(3064..4602)
/gene="TC0442"
/note="similar to GB:M10101 SP:P04079 PID:146276 GB:U00096 PID:1788854; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="GMP synthase"
/protein_id="AAF39296.1"
/db_xref="GI:7190487"
/translation="MSNILIDFGSYTNVLAKIRLLSVFCEVLPWNTPLEKILQKS PGLSIFSGSPHSTVQNNSPEDVKEIYNSNIPILGYCMQIARDFGEVQGGKSEFG YPIFYFPELFGKLQADAFHTEIRMSHDCVVPKDFVIASSQHPPIAISC PK KXLFGLQHPPEVSDSDIGDKILSNFVKHICOTSEWKIEFIEKQLIQIREKVBETE RVALLSGSVDPSSVAVLHNLGDRLSVFDTGLLKKNVEEVKQFSSGLEGELV VDAASKFPDLSGLIEDPEQRKVIGAAFIEVFEASKNLDVQLAQTIIYSDVIESAK SCDAOTIKSHHNVGGPEKLNKLELRLFLFDEVRALGVLPDLVLSRHPFPF PG LGVYVIGVEVRREYVEIVKADSIFIEELKANLYHKVQSAFVLPCKSVAVGDC RHGYTIALRAVESTDPMTACWPSSLREFLNCSRIINEIPEVCVRYVDISDKPPAT IEWE"
complement(4618..5691)
/gene="TC0443"
/complement(4618..5691)
/gene="TC0443"
/note="similar to SP:P06981 GB:X02209 PID:146275 PID:41627 GB:U00096; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="inosine-5'-monophosphate dehydrogenase, putative"
/protein_id="AAF39297.1"
/db_xref="GI:7190488"
/translation="MREALTFDDVLLVPQYSEVLPQDACLTSVSESLSLIPILSAA XDSVTELSMATAMSVAGGIGIVKNDVNAQVAIVKIQSQSTSSVIGVAGIQQGL ERREALVEAGIDTLVVDTHAGHSKLVDLTAFTIKKNPTVLIYGNIVSKAALGLAEI GVDVAVKYGIGSGICTTRIVSGVGLPQLTAMDVAELHDSVRIITADGMRYSGDIV KALAAGACHVLMGLAGTDEAPGEIVQINEHSYKMYRGMSIGAMKGSARVYFOKN NAKKFPVGEVGLVPYKGSILDVLIQLGGIRSGVGLGHNLEELRQNAVFSRITHS GRSSSHIHNLQYIQHAPNYQISK"
6377..6739
/gene="TC0444"
/6377..6739
/gene="TC0444"
/note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF39298.1"
/db_xref="GI:7190489"

```

```

/translation="MFISDKTSLSRHSTDLAFAFGKNLLYPVTOGFGTRISVKPVLYI
DDMTIHGFYSKQSLCKRICCVIACILLAPIVLLMTLIGVIANWCSAIIYKRYQDLYLA
TGLEFGKEMGPWTCAPQI"
gene
complement(6896..7135)
/genes="TC0445"
CDS
complement(6896..7135)
/genes="TC0445"
/translation="MANKIHAITQOIRLORDSCFTEKIQNSSLOPLNVIAGLDELN
PNAEVTGLRSFLPIISAASARVPLNEMNIFSPKTE"
gene
7303..8886
/genes="TC0446"
CDS
7303..8886
/genes="TC0446"
/translation="similar to SP:P26906 PID:48808 GB:AL009126;
identified by sequence similarity; putative"
/genes="TC0447"
/translation="MFKKFLVILIPFLSFTGLTSHOCKEENLRLRVAICHDPMS
LDPROVFLKDVCIKALYEGIVRNDGSLHLALAREXLSODRCVYFFLKKTFWHN
GDLVATYDFEESIKOYFLEHVDNVALRLILAKNSHAVLKGDIPVENLGVRAIDHETL
EITLESFSSFLTITHPVFPVHASLREYIRNRSKSLPIISNGPFIIRCYEPQNL
LIDKNFLYHDQKNSLDVLRQIYVDIHTAVQLFQKYYDVLVGLPWSFFPLEQKVL
SOPKLDYVPLVNLFCNVNHNKPLDNPRLAALSIAIDRETLKLACKGSIATSPVH
PSLSKMPDLVLSQKRIELAKNLAELKVPQBELKTIITLPIESVTLVRAVQEIYR
QQLFDVLGPKISTLGLYHSLFKRSKEPSLSTGNWADYQQAFAFSLILGNCTKK
DYGVDMQNOEYTDIVSRLLIVEDSDILOILASOLLKSLPLIPLYHLDIYAKHPKVS
NLQTSLSGEIDLKRRSLVE"
gene
9068..10144
/genes="TC0447"
CDS
9068..10144
/genes="TC0447"
/translation="similar to GP:2444077; identified by sequence
similarity; putative"
/genes="TC0448"
/translation="MPSPLSTAASSYLRTLQRAFPLGGGYFTPNNSAQTALRVQTA
TSLAIVYPRPLSSPVSEFSKCKNDQAQIENAILSSSVFLKIFSLSSSEPIVD
LICKSESIPVTHIQHIFEPYVKNLEAGAEILRCHNRNRSLLRHKTMILDEKQVIT
GSANFTNSINDVLLARVNVNHNKQKQKGAIIENSTKGTQVLYLINHKK
CGNEROIVKAIKNNTSSQIGMCIITHRGILQALNEATQSVLVITIIIDSLSEQOTI
DILKALGSLRYRVGTGDIRHCKACILDHNTAIGSANWAASGLKANKEDIIVNPLT
EROREDLSIWRYLCDNSAILED"
gene
complement(10175..10582)
/genes="TC0448"
CDS
complement(10175..10582)
/genes="TC0448"
/translation="conserved hypothetical protein"
/genes="TC0449"
/translation="MILKRSYCYLFAWLVSIGITLMSVYVYLLNPEPCVLYYORI
CLPLVVIIGISAYLDDISVKVIALPLALGICIAIYQVQLPEIPGMILDICGVKSCS
TKLFLGLGFTPMASALAFNAIFATKSE"
gene
complement(10382..11295)
/genes="TC0449"
CDS
complement(10382..11295)

```

```

/genes="TC0449"
/translation="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
Query Match 9.0%; Score 161.6; DB 1; Length 11402;
Best Local Similarity 48.0%; Pred. No. 6.3e-29;
Matches 711; Conservative 0; Mismatches 729; Indels 42; Gaps 7;

QY 211 tattaataataagagatgaaccccgcttcttttagatccaagaagtcgagctctttcaga 270
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7407 TGTTGCTATTGGCAGGATCCCAATGCTTTAGATCCTCGCCAAAGTTTTTTGATTAAAGA 7466
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 271 aatcagcccttgtaacacatatatgaagagattagttcaagaagaaataatctttcagaaa 330
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7467 TGTTGTATAGCAAAAACACACTTTATGAAGGATGGTTTCGAGAAAAATGAT-----GSCAG 7520
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 331 tatagagccctgtcttgagaagactactctcttctctcggagcggaactcaactatacttt 390
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7521 TCTTCATTGGCTCTAGCAGAGAAGATACAGCTTGTCAGGATCGGTGTGTCTACATTT 7580
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 391 taaactgaatacagctcttttgagtaatgagcagcccttaacagctgaagactttataga 450
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7581 TTTTCTAAAAAAAACCTTTTGGCATACGGAGATCTTGTAAACAGCATATGATTTTGAAGA 7640
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 451 atcttggaacaaagtagctactcaagaagctctcaggaactctatgcttttgccttgaalcc 510
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7641 GTCTATTAAGCAATTTTATCTCATGAGGTGTGAATGTGTGGCTTACGATTGCTTGGGTT 7700
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 511 aattaaaaattctacgaagatcccaagaggagacacctctccatagaccattttgagtgca 570
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7701 AATTAAAAATCTCATGCTGTCTTCAAAAGGAGACATCTCTGTGTTGAAATTTAGTGTCG 7760
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 571 ctctcctaataatcacacttgcttggtaacccctgggaatcccaacccctgcattcttctaaa 630
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7761 CGCTTTAGATGAACATCTCTTGAATCACTCTAGAACACCCCTCTCTCTCATTTTTTTTGA 7820
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 631 acttttagctcttcagctcttttcccgcttcataaatctcaaaacccctgcataccaa 690
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7821 AACTCTCACATCCGGTCTTTTATCCAGTGCATGCTCTTTTCGAGAGAAATTTATAGAAA 7880
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 691 atc-----tctacctatagcaagcgagctttctctatccctaaaaataca 735
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7881 TCGAAGTAACAGCTCTCTTCCAAATAATTTCTAACGGCCCTTTTATCATTCGGTGTATGA 7940
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 736 acaaaacaaatgataaaaactctcaaaaacccctcaactactataatacaaacagtcagg 795
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7941 GCCACAAAATTTTTTGTGCTGTGATAAAAAATCCTTTTATCATGACCAAAAAATGTTTC 8000
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 796 aactaaaacgattacgattcattcccgatgcaaacacacagcagcaaacactatttaa 855
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8001 TTTAGATGCTGTACGCTTACAAATAGTTCAGGATATTCATACAGGGTGCATTTGTTCA 8060
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 856 tcagggaacaaactcaattggcaggaacctcttggggagaaacgcatctcctcaagaacacct 915
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8061 AAAAAAATATGGAATTTAGTGGGGTTGCCCTTGGAGCTCCTCTCCCTCTAGAGAACA 8120
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 916 atcccaattacagtcctaaagggcgacttctcttcttctgctgcaggaaacctaatgct 975
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8121 AAAAAATCTT---TCTCAGATTTTTTGTATGATATCGGTGCTAAATTCACAGTCTT 8177
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 976 caacttcaataataaattcccccctcaacaatatgaagcttagagaagccttagcctc 1035
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8178 ATTCTGCAATGTTACCAACCAACCATTAGATATCTCTTCGCTAAGACGGCATCTCTTT 8237
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1036 agccttagaagaagcctctgtctcaactatattcttagccgtgcacaaactccga 1095
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8238 AGCAATGATCGAGAAAC-----ATTATTAATATAGCTGCAAGAGCATTTCTAC 8291
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1096 tcatctcctacacaaataattcatagctatcccgaacatcaaaaaaagaagatggcaca 1155
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8292 CAGTTTGTACATCTAGTCTATCTAAGATGCTTTTAGATGTTCTTCGAAAAAAGAGAG 8351
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



/strain="D/UW-3/CX"  
 /db\_xref="taxon:813"  
 206..1384  
 /gene="trpB"  
 /note="Ct170"  
 206..1384  
 /gene="trpB"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Tryptophan Synthase (Beta Chain)"  
 /protein\_id="AAC67761.1"  
 /db\_xref="GI:3328574"  
 /translation="MFKHKHPGGAFLPEELLAPIONLKAWEILLKTOQSFSLSELD  
 LKNGAGROPTLEYKFNKFAKIDGPRVFLKREDLLHTGAHKLNALGOCILAKVLGKT  
 VAAETGAGHGQVATATACAVLGLDGVYVMGAKOVQKPKNEKFEKFGAEWVSYTKG  
 CGKIDVACALQGGNSAWLTHTSFTHYVCIGLSALGPLPYDIPVRFPSVYSIEVKEQIHAVA  
 RPDLLKALIGGSNAPGPHFIPNPKVQLIGVEGGGUGITSSRGHAAEFATGRCGV  
 HGFSYLLQDDGGVQLQTHSI3AGLIDYPSVGPDHMHSSGFAFVTLATDEEALRAFV  
 LTRNEGIIIPALESSHALLALVSIAPSLPKEQIVIVNLSGRCDKLPQIIRNRNGIY  
 \*  
 1377..2138  
 /gene="trpA"  
 /note="Ct171"  
 1377..2138  
 /gene="trpA"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Tryptophan Synthase (Alpha Chain)"  
 /protein\_id="AAC67762.1"  
 /db\_xref="GI:3328575"  
 /translation="MSKLTQVFKFKLCIGYLTAGDGTSTYIEAAKALIQGVDFIL  
 LGFFSPADNPEIQVSHDRALAEINTSETLLEIGVIRAFNOEPLILSYVNPIL  
 QRDLILRLKLDAGIINGVVIDLPAPLSHGKSPFEOLLAVGLDPLILISAGTTPER  
 MSLLQYVARGFLXYINQCAVTRDSEWGIKEEFKRYKREHDFLPVDRDIOCKKEAAHV  
 NYSDGFTVKTAFVHQVTMDSSVEFLTLAQTVIPG"  
 complement(2385..2876)  
 /gene="Ct172"  
 complement(2385..2876)  
 /gene="Ct172"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAC67763.1"  
 /db\_xref="GI:3328576"  
 /translation="MNYENTVTKTMFFELAKRLVOLKNKPNFLKKSETVLFIPEKQ  
 LKMGVYSIDENNYLTDMGSEFGLLOKLANPVCKALQQQVGEAQAATLSNOVTYV  
 DIVLMRSPIFKSVLLETLINEIYQESLFLFKKPNVQCKMSFEHGAHEILLKIFL  
 TVS"  
 complement(2904..3074)  
 /gene="Ct172.1"  
 complement(2904..3074)  
 /gene="Ct172.1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAC67773.1"  
 /db\_xref="GI:3522888"  
 /translation="MLRLFCGLYFPOLPEKOKNLELDIYNKLFEKTVVMGEGACV  
 NLAIGSEFILL"  
 complement(3142..3414)  
 /gene="Ct173"  
 complement(3142..3414)  
 /gene="Ct173"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAC67764.1"  
 /db\_xref="GI:3328577"  
 /translation="MKICREICDALKRDAFTEIRTYKNGSVIYVTSRQCPGVEIRE  
 IGQAKALHNQBELKQAVATUGLKHLLIIPPLSIYKHSLLYKRLVI"  
 complement(3501..3956)  
 /gene="Ct174"  
 complement(3501..3956)





Qy 914 ctatccatttacagtcgttaagggcacttacactcttttgatgctgcaggaacccctcag 973  
 Db 4970 CAAGAATACTCC---CTAGAGAAAAATTATTGATTCCTGCTGATTTGAGTCTCTGTT 5026  
 Qy 974 ctacactcaataataaattccctcccaacaatatgaagcttagaagaccccttagca 1033  
 Db 5027 TTATTCGTACATTCATCAACACCTTTTAAATAATCCCTCGCTGAGAACGCCCTCTCT 5086  
 Qy 1034 tcagccttagaagaagctctgttctcaactatattcttaggcgtgcgcaaaactgcc 1093  
 Db 5087 TTAGCAATCAATCGAGAACTTTATTAAACTAGCAGGTAAGCGCTGACGCTACGAGC 5146  
 Qy 1094 gacatctcttaccatacaataatcttagatgctcccgaaactcaaaacaaagagatgca 1153  
 Db 5147 TTGTGTCACCCAC-----AATTATCTCAGATACCTGCTACTTGTCTCAAGATG--- 5198  
 Qy 1154 caacgcaagcttacgtcaaaaactctttaagaagcttagaagaccccaaatcaact 1213  
 Db 5199 -AGCGATTGCTTTAGCAAAAGGCTACITGACCGAAGCTTTAA-----GACTTTATCT 5251  
 Qy 1214 gctaaagctcgaacatttaattcttctccgcttctccctcgtcagcaagttotta 1273  
 Db 5252 CAAGAAGATTAGAAAAAATACATTAATTTATCTATAGATCTGTTGCTTACGAGCC 5311  
 Qy 1274 ctagtcaaaccttatcagcaaacagtggaagaagtttagggttcgctatccctattgct 1333  
 Db 5312 GTTGTCAAGAAATTCGCCCAACAATATTGATGTAAGTCTGGGATTTAAATAATCTACAITA 5371  
 Qy 1334 ggaagaagattgctctctcctcagcagacactctctcaggaacactctcttagctaca 1393  
 Db 5372 GGAATGATATATCAATGTTTTTTAGCAAAAGCTTCCAGAGGAGAAATCTCTCTTAGCAACT 5431  
 Qy 1394 ggaagatggttcgacagcttgcgtgacccctatgcttccggaacactctcttagctcca 1453  
 Db 5432 GGTAAATGGAATTCGAGACTATCATCAAGCTAGTGCCTTCTGCTGCTAGGTAAATGGG 5491  
 Qy 1454 tcaggagttcctccttatgcatacaaccataaagactctcagaaattctacaaacata 1513  
 Db 5492 ACAAGATATAAAGACTTTCATGTTTAACTGCAGAACCAAAAGTACACAAATATAGTT 5551  
 Qy 1514 gaacagacaaagatcaccacaaacgcctcgggaattagtgctgcagactctcttaccata 1573  
 Db 5552 GCTCAACTCTGATTCAAGAAATCAAGACGACTACAGCTTATGGCAGACAGTGTGCTT 5611  
 Qy 1574 gacactttcatattattgagcagatctaccacagcagcatttcattgctgaataaa 1633  
 Db 5612 AAAGAAAGTCCCTTATTT---CCTCTATACCACTCGATTATGATGATGGAACAGCCT 5668  
 Qy 1634 aaactttcatctagagctctcaccacacagaggttggtgactt 1677  
 Db 5669 CGGGTCTGATCTCCAAACCTCTCTCTGCTGGAGAAATGATTT 5712

RESULT 15  
 AF305387  
 LOCUS  
 DEFINITION Bacillus thuringiensis 9364 bp DNA linear BCT 30-AUG-2001  
 sequence.  
 ACCESSION AF305387  
 VERSION AF305387.1 GI:10717142  
 KEYWORDS  
 SOURCE Bacillus thuringiensis.  
 ORGANISM Bacillus thuringiensis.  
 Bacteria; Firmicutes; Bacillus/clostridium group;  
 Bacillus/Staphylococcus group; Bacillus; Bacillus cereus group.  
 1 (bases 1 to 9364)  
 Gominet,M., Slamti,L., Gillois,N., Rose,M. and Lereclus,D.  
 Oligopeptide permease is required for expression of the Bacillus  
 thuringiensis plcR regulon and for virulence  
 JOURNAL Mol. Microbiol. 40 (4), 963-975 (2001)  
 MEDLINE 21295338  
 PUBMED 11401703  
 REFERENCE 2 (bases 1 to 9364)

AUTHORS  
 TITLE  
 JOURNAL

FEATURES  
 source

Gominet,M., Gillois,N., Rose,M. and Lereclus,D.  
 Direct Submission  
 Submitted (13-SEP-2000) Biotechnologies, Institut Pasteur, 25 rue  
 du Docteur Roux, Paris 75015, France  
 Location/Qualifiers  
 1. 9364  
 /organism="Bacillus thuringiensis"  
 /db\_xref="taxon:1428"  
 1249. 7088  
 /gene="oligopeptide permease operon"  
 1332. 2987  
 /gene="oppA"  
 1332. 2987  
 /gene="oppA"  
 /codon\_start=1  
 /transl\_table=11  
 /product="OppA"  
 /protein\_id="AAG22034.1"  
 /db\_xref="GI:10717143"  
 /translation="MKKKIPILLASTLTASMLLAGCSYQKDDAKAKKENTSSNGK  
 QVNLDELSEIPSDASLASDSASTALNTMEGLTRIGKDKRVPVGIADVDQKDDG  
 KKYIFKLKRDARWSNGBEVTAQDFVYSKRAVNPDTKATSYIMFDIKNAEKIHKKE  
 PADQGVKALDDITILEVDLPVDFYVPLNENFVKQAQDQGLKLEANTL  
 YNGPFVDMKHEQSFQKKNPSYDNKTKYIEENFNVTSTDNVLYETSDIRA  
 ALTSFVDFKRSSEFTRKEAGVAVLRFNOSNYLSNKLKALSNKSDPDNIKVI  
 LNNGATGAYFVGKDFAGNKKDFRAENGKLVETNPKEAKLWETAKKELGDKTEL  
 EFLPNDEDAKVGFEKLGEMKNLPGLSIKIQQPFQAKNNKLEDSQQYDIAGFWGP  
 DRPDTSYLDMPVTNGSNKYDELILKARTDKDLQARNNLLVEKMLIK  
 EDVATIPFQKGSAYIVKDAVKDILIPINTGGRLTYKWAQVEQ"  
 3115. 4044  
 /gene="oppB"  
 3115. 4044  
 /gene="oppB"  
 /codon\_start=1  
 /transl\_table=11  
 /product="OppB"  
 /protein\_id="AAG22035.1"  
 /db\_xref="GI:10717144"  
 /translation="MGRYVLRKRVYVMAALTFLITLTFILMKLLPGSLKNOEKLSPA  
 QKEITELKYLNDPVPVQVARYLGNLAKGDLGVSFOYDNPVTDIVDRIGFSAQLG  
 PAAILGLVFIILGIVAAALNNNTWDYCATILISVIGMSVSPFVAALQYFVGVKLG  
 PVAFWKPGETVNPVIALSMAVIATIAFRAELIEVMQADYILTAKAGISQGVII  
 VKHALRNALIPVVTILGPMVAALITGTLVIEQIYAVPGLGEQFVKSITVNDYTVIMGT  
 TIFSAIFILVIFIVDILYGIIDPRIRLAGGK"  
 4044. 5057  
 /gene="oppC"  
 4044. 5057  
 /gene="oppC"  
 /codon\_start=1  
 /transl\_table=11  
 /product="OppC"  
 /protein\_id="AAG22036.1"  
 /db\_xref="GI:10717145"  
 /translation="MKDYKLSPLDFOAQNQNVNVEIARSLTFWQDVRRLFOHK  
 GAMTGVLLSLIILAIKFGPVWSKSHKEDLGRKMPKIPVENVHWPFGTQY  
 GVQYKEDIKKEFFNFQDDLDRLTWRTWESTVSLYALAAALDVLGVAYGGS  
 AFVGRVDMIMORIMEIINGIPYLIIVLIMVINGSIWSTLAMAITGTMGRIVR  
 GQILKNOEYVLAISRTGLATQILVHLIPNMGPITVMTPTIPSAVGEAFSLF  
 IGLGIQFPFASISLVNDGYSIQTYPHMFIIPAVVISILILANLMDGLRDLDPK  
 MRK"  
 5076. 6119  
 /gene="oppD"  
 5076. 6119  
 /gene="oppD"  
 /codon\_start=1  
 /transl\_table=11  
 /product="OppD"  
 /protein\_id="AAG22037.1"  
 /db\_xref="GI:10717146"  
 /translation="MKTLLKVKQLQVSDTHAGEVQAVRGVTFDLKGETLAIVGSSG  
 SGKSVTSKALMGLIPNPPGRIKNGELVFEGRDLTKLEKEMQVGRKQIAMIFQDPM  
 SLNPTWTIGNOIMEGLIKHOGSKADARKVALELIDLVGINPPEARLKQYHQFSGGM  
 RQVVVIAALACNFKLLIADEPTALDVTIQAGIILELMKDIOQKTEAAIFITHDLGV

gene  
 CDS

gene  
 CDS



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 26, 2002, 03:00:52 ; Search time 299.98 Seconds  
(without alignments)  
10296.450 Million cell updates/sec

Title: US-09-824-567-1

Perfect score: 1799

Sequence: 1 acttccccctgctaaacta.....tgactgctctctctctcta 1799

Scoring table: IDENTIFY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1799	100.0	1799	22	AAD20238
2	1786.4	99.3	1230025	20	AAX91990
3	449.6	25.0	1038602	20	AA201425
4	255	14.2	1235	24	AA518752
5	181.4	10.1	1787	24	AA518750
6	74.6	4.1	1536	20	AAX61764
7	74.6	4.1	1587	20	AAX61763
8	63.8	3.5	1028	20	AA20051
9	63.8	3.5	1121	20	AA20051

10	63.8	3.5	1122	20	AAX20050	Enterococcus faeca
11	63.6	3.5	1564	20	AA51864	B. subtilis opp op
12	62.6	3.5	1564	20	AA20031	Enterococcus faeca
13	62.6	3.5	1641	20	AA20030	Enterococcus faeca
14	62.6	3.5	9021	20	AA20032	Enterococcus faeca
15	60.4	3.4	15747	20	AA20031	Enterococcus faeca
16	59.6	3.3	1521	20	AA20031	B. burgdorferi ant
17	59.6	3.3	1572	20	AA20031	B. burgdorferi ant
18	57.6	3.2	1038602	20	AA201425	Complete genome se
19	55	3.1	1579	20	AA20041	Enterococcus faeca
20	55	3.1	1680	20	AA20040	Enterococcus faeca
21	55	3.1	5504	20	AA20040	Enterococcus faeca
22	52.2	2.9	26811	20	AA20253	Borrelia burgdorfe
23	50.8	2.8	716	21	AA63289	C. trachomatis LGV
24	50.8	2.8	716	22	AA456193	Chlamydia trachoma
25	49.6	2.8	1597	20	AA20043	Enterococcus faeca
26	49.6	2.8	1680	20	AA20042	Enterococcus faeca
27	49.6	2.8	9509	20	AA20042	Enterococcus faeca
28	48	2.7	1653	23	AA52765	Enterococcus faeca
29	47.6	2.6	6290	24	ABL33046	Human immune syste
30	47.2	2.6	1545	20	AA61686	B. burgdorferi ant
31	47.2	2.6	1626	20	AA61685	B. burgdorferi ant
32	46	2.6	10996	20	AA20057	Human immune syste
33	46	2.6	18997	24	ABL32570	Enterococcus faeca
34	45.6	2.5	1585	20	AA20021	Enterococcus faeca
35	45.6	2.5	1659	20	AA20020	Enterococcus faeca
36	45.6	2.5	11739	20	AA20020	Enterococcus faeca
37	44.8	2.5	1119	20	AA20057	Enterococcus faeca
38	44.8	2.5	1242	20	AA20056	Enterococcus faeca
39	44.8	2.5	2991	20	AA20056	Enterococcus faeca
40	44.4	2.5	16842	22	AA46411	Enterococcus faeca
41	44.4	2.5	16842	22	AA46411	Enterococcus faeca
42	44.2	2.5	747	20	AA20057	Tumour suppressor
43	44.2	2.5	7309	22	AA20057	Nucleic acid sequ
44	44.2	2.5	7309	22	AA20057	Tumour suppressor
45	43.2	2.4	6200	22	AA46442	Human immune syste

#### ALIGNMENTS

#### RESULT 1

AAD20238  
ID AAD20238 standard; DNA; 1799 BP.

XX AC AAD20238;

XX AC AAD20238;

XX AC AAD20238;

XX AC AAD20238;

XX AC AAD20238;

XX AC AAD20238;

XX AC AAD20238;

XX AC AAD20238;

XX AC AAD20238;

XX AC AAD20238;

XX AC AAD20238;

XX AC AAD20238;

XX AC AAD20238;

XX AC AAD20238;

XX AC AAD20238;

XX AC AAD20238;

XX AC AAD20238;

XX AC AAD20238;

XX AC AAD20238;

DR WPI: 2001-648549/74.  
 DR P-PSDB; AAE12212.  
 XX  
 PT Novel Chlamydia ATP-binding cassette and corresponding DNA molecule for  
 PT preventing, diagnosing and treating Chlamydia infections in mammals, in  
 PT particular humans -  
 XX  
 PS Claim 2; Fig 1; 80pp; English.  
 XX  
 CC The present invention relates to novel Chlamydia pneumoniae ATP-binding  
 CC cassette protein and its corresponding gene. Sequences of the invention  
 CC are useful for detecting Chlamydia infection by assaying a body fluid  
 CC of a mammal with the components. They are also used as vaccines. ATP  
 CC binding cassette antibodies and vaccines of the invention are useful  
 CC for preventing or treating Chlamydia infection e.g. infection caused  
 CC by C. trachomatis, C. psittaci, C. pneumoniae or C. pecorum in mammals,  
 CC such as humans. The nucleic acid molecules are useful for producing  
 CC ATP-binding cassettes, in the construction of vaccine vectors such  
 CC as poxviruses, which are further useful for preventing and/or treating  
 CC Chlamydia infection and in the construction of attenuated Chlamydia  
 CC strains that can over-express the nucleic acid molecules or express  
 CC it in a non-toxic, mutated form. The present sequence is a gene encoding  
 CC Chlamydia pneumoniae ATP-binding cassette.  
 XX  
 SQ Sequence 1799 BP; 560 A; 439 C; 294 G; 506 T; 0 other;

Query Match 100.0%; Score 1799; DB 22; Length 1799;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1799; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acttccccctgtataactatgctcagataatgctgtatgattgaggtctaggggag 60  
 Db 1 acttccccctgtataactatgctcagataatgctgtatgattgaggtctaggggag 60  
 QY 61 aaaaatttccaaaaaacttagtattccggaattctgtatgctcgaagataatcagttg 120  
 Db 61 aaaaatttccaaaaaacttagtattccggaattctgtatgctcgaagataatcagttg 120  
 QY 121 aatctgtatcacattctctttagctctccgttagtctccgaaggtcgaaggtccag 180  
 Db 121 aatctgtatcacattctctttagctctccgttagtctccgaaggtcgaaggtccag 180  
 QY 181 tcaactctctacatctcggggagactcgtattataataagagatgaaccccttttt 240  
 Db 181 tcaactctctacatctcggggagactcgtattataataagagatgaaccccttttt 240  
 QY 241 agatccagaagaagtcgactcttctcagaatcagcttgctcgaaggtcgaaggtc 300  
 Db 241 agatccagaagaagtcgactcttctcagaatcagcttgctcgaaggtcgaaggtc 300  
 QY 301 attagttcagaagaataattcttcagggaatatagagctgctcttcgagaagactctc 360  
 Db 301 attagttcagaagaataattcttcagggaatatagagctgctcttcgagaagactctc 360  
 QY 361 tcttctctcggagactcaactatacttttaaacgaatacagcttttttggagtaatgg 420  
 Db 361 tcttctctcggagactcaactatacttttaaacgaatacagcttttttggagtaatgg 420  
 QY 421 cgaaccttcaacgctgaagactttatagaattcttggaacagtagctactcaagaagt 480  
 Db 421 cgaaccttcaacgctgaagactttatagaattcttggaacagtagctactcaagaagt 480  
 QY 481 ctcagggaatctatgcttttgccttgaatccaattaaaaatgtacgaagaatccaagggg 540  
 Db 481 ctcagggaatctatgcttttgccttgaatccaattaaaaatgtacgaagaatccaagggg 540  
 QY 541 acaactctcatagaccattttggagtcgactctcotaatgaatcactatgtttgttac 600  
 Db 541 acaactctcatagaccattttggagtcgactctcotaatgaatcactatgtttgttac 600  
 QY 601 cctggaatcccccaactcgcattttcttaaaacttttagctcttcagctcttttccccct 660  
 Db 601 cctggaatcccccaactcgcattttcttaaaacttttagctcttcagctcttttccccct 660

Db 601 cctggaatcccccaactcgcattttcttaaaacttttagctcttcagctcttttccccct 660  
 QY 661 tcaataatctcaaaagaacccctgcaatccaatctctacatctacatagcaagcgactttcta 720  
 Db 661 tcaataatctcaaaagaacccctgcaatccaatctctacatctacatagcaagcgactttcta 720  
 QY 721 tcttaaaaaatcaaaacaaatgagataaaactctcaaaaaactcactactataa 780  
 Db 721 tcttaaaaaatcaaaacaaatgagataaaactctcaaaaaactcactactataa 780  
 QY 781 tcaaatcgagggtggaactaaacgattacgattcacttccatcccgatgcaaacacagc 840  
 Db 781 tcaaatcgagggtggaactaaacgattacgattcacttccatcccgatgcaaacacagc 840  
 QY 841 agcaaaactatttaatacagggaacaaactcaattggcaaggacctctctggggaagcgaat 900  
 Db 841 agcaaaactatttaatacagggaacaaactcaattggcaaggacctctctggggaagcgaat 900  
 QY 901 tctcagaagaacctatccaattacagttcaggtcgaaggtcacttcaactcttttgaatgcg 960  
 Db 901 tctcagaagaacctatccaattacagttcaggtcgaaggtcacttcaactcttttgaatgcg 960  
 QY 961 aggaacctcaggtcgcacccctcaatatacaataaaattccccctcaacaatatgaagcttag 1020  
 Db 961 aggaacctcaggtcgcacccctcaatatacaataaaattccccctcaacaatatgaagcttag 1020  
 QY 1021 aggaacctcaggtcgcacccctcaatatacaataaaattccccctcaacaatatgaagcttag 1080  
 Db 1021 aggaacctcaggtcgcacccctcaatatacaataaaattccccctcaacaatatgaagcttag 1080  
 QY 1081 tgcataaaactcgcgactcattctcactcaacaatatcattcatagctatcccgaaacataaa 1140  
 Db 1081 tgcataaaactcgcgactcattctcactcaacaatatcattcatagctatcccgaaacataaa 1140  
 QY 1141 acaagagatggaacacgcgcaagcttacgtaacaaactctttaaagaagctttagaaga 1200  
 Db 1141 acaagagatggaacacgcgcaagcttacgtaacaaactctttaaagaagctttagaaga 1200  
 QY 1201 actccaaactcactcgtcaagatctcgaacatcttaattcttctcccggttctctcgctc 1260  
 Db 1201 actccaaactcactcgtcaagatctcgaacatcttaattcttctcccggttctctcgctc 1260  
 QY 1261 agcaagttctttagctccaaacttatacgaagacagtggaagaagtttaggggttcgc 1320  
 Db 1261 agcaagttctttagctccaaacttatacgaagacagtggaagaagtttaggggttcgc 1320  
 QY 1321 tatccctatttgcggaagaaatttgcctcttcccaagcagacactatcttcagggaactt 1380  
 Db 1321 tatccctatttgcggaagaaatttgcctcttcccaagcagacactatcttcagggaactt 1380  
 QY 1381 ctctttagctacagagaggtggttcgagacttttgcgtatccttgcgtatggcattctaaagat 1440  
 Db 1381 ctctttagctacagagaggtggttcgagacttttgcgtatccttgcgtatggcattctaaagat 1440  
 QY 1441 ctctttagctacagagaggtggttcgagacttttgcgtatccttgcgtatggcattctaaagat 1500  
 Db 1441 ctctttagctacagagaggtggttcgagacttttgcgtatccttgcgtatggcattctaaagat 1500  
 QY 1501 tctcaaaactatagaacagagcgaagatacaaaaaactcgcgaattagttgcgaagc 1560  
 Db 1501 tctcaaaactatagaacagagcgaagatacaaaaaactcgcgaattagttgcgaagc 1560  
 QY 1561 tctctttttagctagagacttttcaattattattgagcagatctaccacagcagacttcaatt 1620  
 Db 1561 tctctttttagctagagacttttcaattattattgagcagatctaccacagcagacttcaatt 1620  
 QY 1621 tgcgtatgaataaaaaacttttcaattctagggagttccaccaacagaggttgcgacttcgc 1680  
 Db 1621 tgcgtatgaataaaaaacttttcaattctagggagttccaccaacagaggttgcgacttcgc 1680  
 QY 1681 ttatgctaaagaaatttagcacctcttttaattctcgaacttttgaacttgcgaagactgaatt 1740  
 Db 1681 ttatgctaaagaaatttagcacctcttttaattctcgaacttttgaacttgcgaagactgaatt 1740

Db	244560	tcactctctcaactctcgggggaactcgtctattataaagagatgaacccctctttt	244661
QY	241	agatccaagaacagtgagactcttttcagaaatcagccttgccaacatctctatgagg	300
Db	244620	agatccaagaacagtgagactcttttcagaaatcagccttgccaacatctctatgagg	244679
QY	301	attagtttcaagaaataatcttttcaggaaatatagagctgctgttcgagaagactactc	360
Db	244680	attagtttcaagaaataatcttttcaggaaatatagagctgctgttcgagaagactactc	244739
QY	361	tctttctcggcagcgactcaactatacttttaaacTgaaatcagcttttttgagagtaatgg	420
Db	244740	tctttctcggcagcgactcaactatacttttaaacTgaaatcagcttttttgagagtaatgg	244799
QY	421	cgaccccttaacagctgagacctttatagaaatbtggaacagtagctactcaagaagt	480
Db	244800	cgaccccttaacagctgagacctttatagaaatbtggaacagtagctactcaagaagt	244859
QY	481	ctcaggaaatctatgcttttgcttTgaaatccaattaaaaatTgaaagatcccaagagg	540
Db	244860	ctcaggaaatctatgcttttgcttTgaaatccaattaaaaatTgaaagatcccaagagg	244919
QY	541	acacctctcaatagaccattttggatTgcactctctaaTgaatctacactTgtttttac	600
Db	244920	acacctctcaatagaccattttggatTgcactctctaaTgaatctacactTgtttttac	244979
QY	601	ccTggaatcccaacctcgcatctttTaaacttttttagctcttccagcttttttcccggt	660
Db	244980	ccTggaatcccaacctcgcatctttTaaacttttttagctcttccagcttttttcccggt	245039
QY	661	tcataactcTcaagaaacctTgcataccaactctctacactatagaaTgcgagcttttcta	720
Db	245040	tcataactcTcaagaaacctTgcataccaactctctacactatagaaTgcgagcttttcta	245099
QY	721	tcctaaaaatTcaacaaaaaacatTgataaaaactctcaaaaaacctcactactataa	780
Db	245100	tcctaaaaatTcaacaaaaaacatTgataaaaactctcaaaaaacctcactactataa	245159
QY	781	tcaaaTcaggTggaaactaaaacgattacgattcacttcacttcctcagTcaaacacagc	840
Db	245160	tcaaaTcaggTggaaactaaaacgattacgattcacttcacttcctcagTcaaacacagc	245219
QY	841	agcaaaactatttaactcagggaaacatccaattTgcaaggaacctctTggggaagacgat	900
Db	245220	agcaaaactatttaactcagggaaacatccaattTgcaaggaacctctTggggaagacgat	245279
QY	901	tcctcaagaaacctatccaattTcagTctaaggggcactTcaactctttttgatTgcg	960
Db	245280	tcctcaagaaacctatccaattTcagTctaaggggcactTcaactctttttgatTgcg	245339
QY	961	aggaacctcagTgcaccttcaatatccaataaatttccccctcaacaatatgaagcttag	1020
Db	245340	aggaacctcagTgcaccttcaatatccaataaatttccccctcaacaatatgaagcttag	245399
QY	1021	agaagccttagcatcagccttagataaggaagcctctTgtccaactatatcttttaggcg	1080
Db	245400	agaagccttagcatcagccttagataaggaagcctctTgtccaactatatcttttaggcg	245459
QY	1081	Tgcaaaaactcgcgatcatctcactacaataattcaatagctctccggaacatcaaaa	1140
Db	245460	Tgcaaaaactcgcgatcatctcactacaataattcaatagctctccggaacatcaaaa	245519
QY	1141	acaagagatggcacaacgcgaagcttagctTaaaaaaactctttTaaagaagctttagaaga	1200
Db	245520	acaagagatggcacaacgcgaagcttagctTaaaaaaactctttTaaagaagctttagaaga	245579
QY	1201	actccaaatcaactTgctaaagatctcTgaaactTtaattcttatcttttcccgTttccctgc	1260
Db	245580	actccaaatcaacgcTaaagatctcTgaaactTtaattcttatcttttcccgTttccctgc	245639
QY	1261	agcaagttctttactagTccaacttatcagagaacagTggaaagaagtttaggttcgc	1320
Db	245640	agcaagttctttactagTccaacttatcagagaacagTggaaagaagtttaggttcgc	245699



QY 1321 tatccctattgttggaaggaattgtctcttcccaagcagacctatcttcaggaactt 1380  
 Db 245700 tatccctattgttggaaggaattgtctcttcccaagcagacctatcttcaggaactt 245759  
 QY 1361 ctctttagctacagagagatggttcgagactttgtctgctatgctatgctatttaacgat 1440  
 Db 245760 ctctttagctacagagagatggttcgagactttgtctgctatgctatgctatttaacgat 245819  
 QY 1441 ctctgtctatccatcagagagttctctctatgcaatcaacataagagacttcagaaat 1500  
 Db 245820 ctctgtctatccatcagagagttctctctatgcaatcaacataagagacttcagaaat 245879  
 QY 1501 tctacaaacatagacaagagcaagatcaccaaaacgctcggaattagtgctgcaagc 1560  
 Db 245880 tctacaaacatagacaagagcaagatcaccaaaacgctcggaattagtgctgcaagc 245939  
 QY 1561 tctctttacctagagacctctctatattattagccgacttaccacagcgcattcaatt 1620  
 Db 245940 tctctttacctagagacctctctatattattagccgacttaccacagcgcattcaatt 245999  
 QY 1621 tctatgaataaaaaactttc-taotcagagagctctcaccacacagaggttgctgacttcc 1679  
 Db 246000 tctatgaataaaaaactttc-taotcagagagctctcaccacacagaggttgctgacttcc 246059  
 QY 1680 gttatcctaaggaataattagcacctcttttaattctctgcaaaacttgctcaagaactgaatct 1739  
 Db 246060 gttatcctaaggaataattagcacctcttttaattctctgcaaaacttgctcaagaactgaatct 246119  
 QY 1740 tatactaaactggttgctcttggtgacactctgttctctctgactctctctctcteta 1799  
 Db 246120 tatactaaactggttgctcttggtgacactctgttctctctgactctctctctcteta 246179

RESULT 3  
 AAZ01425/C  
 ID AAZ01425 standard; DNA; I038602 BP.

XX AC AAZ01425;

XX DT 07-OCT-1999 (first entry)

XX DE Complete genome sequence of Chlamydia trachomatis.

XX KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW Bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.

XX OS Chlamydia trachomatis.

XX PN WO9928475-A2.

XX PD 10-JUN-1999.

XX PF 27-NOV-1998; 98WO-IB01939.

XX PR 04-NOV-1998; 98US-0107077.

XX PR 28-NOV-1997; 97FR-0015041.

XX PR 17-DEC-1997; 97FR-0016034.

XX PA (GEST ) GENSET.

XX PI Griffais R;

XX DR WPI; 1999-371125/31.

XX PT Genome sequence of Chlamydia trachomatis

XX PS Claim 1; Page 373-656; 1755pp; English.

XX CC The present sequence represents the complete genome of Chlamydia

trachomatis. Open reading frames (ORFs) of the genome encode

CC polypeptides AAV36754-Y37949. The polypeptides can be used as vaccines  
 CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also  
 CC be used to control growth of the microorganism. Chlamydia trachomatis is  
 CC responsible for a large number of diseases, e.g. eye diseases such as  
 CC conjunctivitis; genital diseases such as nongonococcal urethritis;  
 CC epididymitis, cervicitis, salpingitis, perihhepatitis, Bartholinitis;  
 CC pneumopathy in breast feeding infants; and venereal  
 CC lymphogranulomatosis. The polypeptides of the invention may be of use in  
 CC treating these diseases.

XX  
 SQ Sequence I038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

Query Match 25.0%; Score 449.6; DB 20; Length 1038602;  
 Best Local Similarity 56.7%; Pred. No. 2.4e-115;  
 Matches 976; Conservative 0; Mismatches 696; Indels 50; Gaps 6;

QY 1 acttccccctgctaaactatgctcagataatgctgctatgattgaggtctaggggag 60  
 Db 252200 ACTTCCCTTCTTCGCGAGTTATGTTTCAGACAACGCTGCAATGATAGTGGTTAGGAGAGA 252141  
 QY 61 aaaaattttcaaaaaaactctagattcc---ggaaattgttatatgctgcaagatatcagt 117  
 Db 252140 GACTGTTTTCGAATCGCACACATGTTTCTAAGGAGGTCATTCCATGCGCAAGATATCAGT 252081  
 QY 118 ggggaattgtatcacattctcttagctctcgttagtctctccaaggctcaaggagtc 177  
 Db 252080 GGGAAATCTGTTCTGCTTCTAGCATAGCAACTTCTGGATGTTCAAAATCTCTCTTAACGC 252021  
 QY 178 cagtcactctctacatctccgggagaaactcgtctattataataagagatgaacccgttc 237  
 Db 252020 AACCCATCGGCTCTCAGCTACTCACACAGTTGCTGTGAACGTAAGAAAGATGATCTCGCAC 251961  
 QY 238 tttagatccaagaacaagtgcgactcttttcagaaatcagcttgcacacatatctatga 297  
 Db 251960 ATTGTATCTTCGAGAGGTTGCGCTTCTTCTGATATCAATTTGATTCATCATCTCTATGA 251901  
 QY 298 gggattagttccaagaaaaataatcttctcaggaaatatagagctctcttgcagaagacta 357  
 Db 251900 AGGATTGGTACAGAA---ACTCCTCTTGGAGAAAGCTCTTCCCTGCTTTAGCGGAGAGTTT 251844  
 QY 358 ctctcttctctcgcgcgactcaacttatacttttaactgaaatcagctcttttttgagtaa 417  
 Db 251843 CTTCCTTATCCGAGATATAAAACTTATACCTTCACTTCACTTCACTTCACTTCACTTCACTT 251784  
 QY 418 tgcgaccccttaacagctgaagactttatagaactcttggaaacagtagctactcaaga 477  
 Db 251783 TGGAGATCTTATTACCGCTCATGATTTTGTTCGTTCTGGAATGATGTTTACAAATCG 251724  
 QY 478 agtctcaggaatcagctgttctgcttgaatccaattaaaaatgcaagaatccaaga 537  
 Db 251723 TGTGCGTAGTATTATTCTTTCGCTTCTTCCTCTATGA----- 251683  
 QY 538 gggacacctctccatagaccatttttgagtgcaactctcttaactgaatcactcttct 597  
 Db 251684 -----CGTGAATAGGATCTGGATTTTSCCAAGATGATCATCTCTTGTAT 251634  
 QY 598 taccctggaatcccaacctgattcttaaaacttttagctctctcagctcttttcc 657  
 Db 251633 CAATCTCTCTACTCCCACTCCACATTTTCTAAAGCTGTTTACCCTCCCTATTTATCC 251574  
 QY 658 cgttcat---aaatccaagaacctgcaatccaatctctactactagcagagagc 714  
 Db 251573 TGTGCAATTCGAGCATGATATGGAAGAAAGAAATCTCTCCGATATCTACTGGAGC 251514  
 QY 715 ttctatctctaaaaatatcaaaaaaacaatgagataaaactctcaaaaaacctcaacta 774  
 Db 251513 TTTTCTTAAAGAGAGAGAAAGACCCGAAGATGGTTAAGCTAGAGAAGAGCCCTTACTA 251454  
 QY 775 ctataatcaagtcagtgaggaaactaaacagattacgattcacttccatcccgatgcaaa 834  
 Db 251453 CTATAATAAGAACCCAGGTAGCTGTACAGGAGATCTGTATACACATCATCTCTGATACACA 251394





QY	262	tcttcagaaatcagccctgttccaaactatctatgagggattagtttccagaaataatctc	331
DB	271	aagcagagatgttccctcagaaaaagccctctatgaaggatgacaagagaactgac-	329
QY	322	tccaggaaataagagccctgcctcttgacagaagactactcttctccggcggactcac	381
DB	330	-----aaggaatcgactgctcttgagaagtataccctgtcacaagatcataaggt	384
QY	382	ttatactttaaactgaaatcagctcttttgagtaatggcagcccttaacagctgaaga	441
DB	385	ctatacctttaaactcagacctctctgtggagcgatggcaactccactcaactgatga	444
QY	442	ctttatagaactcttgaaacaagtagtactcaagaagctcaggaatctatgctttgc	501
DB	445	ctttgaaaactataaaacaacgtacttcgagaattttcaaccttcatacatacttt	504
QY	502	cttgaatccaattaaaaatgtacagaagatccaagaaggaggaacacttcctagaccatt	561
DB	505	actcggcgtgatataaaattcttcggcaatcccaactgctcaaaaatctcttgaaaactc	564
QY	562	tgagtgcaactctcctaataagaatcacacttgtgtaccctggagatccccaactcgca	621
DB	565	tgggatacaggcaaaagatgatcttactctgggataccctcagagcaacttcccata	624
QY	622	tttctaaacttttagctcttccagctcttttcccctgctcata-----aatc	669
DB	625	ctttctcaacttatcgctgcgccgtattctcccctgttccatcacaccttagggaac	684
QY	670	tcaagaacctgcataccaatctctaacatatagacaaggcggagcttctatcctaaaa	729
DB	685	ctataagaagggaacaccccatcccatcacatctccaatgggccttctgtctaaaaaa	744
QY	730	tatcaacaacaaacaaatggataaaactctcaaaaaacccctcaactataatcaaatca	789
DB	745	acatgaacacacaaacctacttaatttagaaaaaatctccactatacttgatcgaac	804
QY	790	ggtggaactataaacgattacgattcaactctccgaatcccgatgcacaacagcagcaaat	849
DB	805	agtaagtgtagccggagtcacacttaaaaattatccagagcgctccacagccagaaact	864
QY	850	atttaatcagggaacctcaattggcaaggacctccttggggagaacgacttcctcaaga	909
DB	865	tttcaagaataactatagatggattggctcactctggagcgctccgatatactaacga	924
QY	910	aacctctcaaatttacagcttaaggggcaactcaactcttttgatctgcaggaacctc	969
DB	925	agacca---aagttctctcccaagaagaattcttaacctatctgttccaagccacc	981
QY	970	atggctcaccttcaatatcaataaattccccctcaacaatatgaagcttagagaacctt	1029
DB	982	ccctctctatctaaccctgcacaaaccttaatacaaaaataaagccctcaggaaagccat	1041
QY	1030	agcatcagccttagataagggaagctctgtctcaactatattcttaggcctgcacaaac	1099
DB	1042	tgctcagctattgtatagaaattctatgaagctcgtgccttcaggacaagaa-----	1096
QY	1090	tgccgatctactcctcaactacaataatcaatgcattcccgacaactcaaaaacaagat	1149
DB	1097	-gcgtgaactctagttcccccatacttttcaactcaactcaacttcaaaagagatctcaac	1155
QY	1150	ggcaacagccagcttagctctaaaaaccttttaagaagcttttagagaactcccaat	1209
DB	1156	agaagaagcacaacaaagccagagcatattttcaagaagcttaagaagaacac-----	1209
QY	1210	cactgctcaagatctcgaacatctaatcttatctttccgctttcctgcgaagcttc	1269
DB	1210	ttctgaaaaagacctcgcagaactcgaactcgaactcctctcctatagattccctcgaattc	1269
QY	1270	tttactagtcacacttatcagagaacagtggaagaagaagtttagggttcgctatccat	1329
DB	1270	catcatagctcaagaataccaagacaacttaaaatgaatacttaggattgaaaaacaaat	1329
QY	1330	tgctcgaaaggaaatttgctcttctccaagcagacactatctccaaggaaactctcttagc	1389







Qy	371	gacgactcacttatacttttaaaactgaaatcagcttttttggagtaagtgcgacccctta	433	
Db	280	gacggttagctatatactcaagtacgagaaagcaaaatggcaaacgggagtcagtt	339	
Qy	431	acagctgaagactttatagaattcttggaaacaagttagctactcaagaagtctcaggaa	490	
Db	340	aaagcaggggattttgtagtcgcttttagaaacgtgggtcgatccagcatcaggttcaagt	399	
Qy	491	tatgcctttgcctgaatccaattaaaaatgtacgaaagatccaagagggacacctctcc	550	
Db	400	agcadgaatcaaatggatatttttaaaatggcgcgcggtgcggagggacagccagc	459	
Qy	551	atagaccattttggagtgacactctcctaatgaatctacactttgttacactggaatcc	610	
Db	460	atggaagaatttggtgtcaagcaatcgatgaccagacactagaaactaacttggaat	519	
Qy	611	ccaactcgcattctttaaacttttagctcttcacgtctcttttcccgcgttcataatc	669	
Db	520	ccaattcttatttgcccaagtctggttgggacacctttatgctcaaaatgaagc	578	
RESULT 11				
AAAS1864				
ID	AAAS1864 standard; DNA; 6300 BP.			
XX				
AC	AAAS1864;			
DT	31-OCT-2000 (first entry)			
XX				
DE	B. subtilis opp operon.			
XX				
KW	Opp operon; SpOok; oligopeptide permease; sporulation; ABC transporter;		ATP-binding cassette transporter; mutation; protein production; oppa; oppb; oppc; oppd; oppF; ds.	
KW	ATP-binding cassette transporter; mutation; protein production; oppa;			
KW	oppb; oppc; oppd; oppF; ds.			
OS	Bacillus subtilis.			
XX				
Key	Location/Qualifiers			
FT	CDS	477..2110	/*tag= a /label= oppA /product= ligand-binding_protein /transl_except= (pos:2106..2107,aa:Phe) /note= "this codon contains an apparent 1 base deletion which alters the reading frame"  2222..3157 /*tag= b /label= oppB /product= membrane_protein 3161..4078 /*tag= c /label= oppC /product= membrane_protein 4083..5159 /*tag= d /label= oppD /product= Atpase 5152..6078 /*tag= e /label= oppF /product= Atpase	
FT				
FT				
FT				
FT				
FT				
FT				
FT				
FT				
FT				
FT				
FT	CDS			
FT				
FT				
FT				
FT				
FT				
FT				
FT				
FT				
FT				
FT	CDS			
FT				
FT				
FT				
FT				
FT				
FT				
FT				
FT				
FT				
FT	CDS			
FT				
FT				
FT				
FT				
FT				
FT				
FT				
FT				
FT				
FT	CDS			
FT				
FT				
FT				
FT				
FT				
FT				
FT				
FT				
FT				
PN	W0200039323-A2.			
XX				
PD	06-JUL-2000.			
XX				
PF	21-DEC-1999; 99WO-US31010.			
XX				
PR	24-DEC-1998; 98GB-0028711.			
XX				
(GENM )	GENENCOR INT INC.			
PA	(DIAG2) DIAZ-TORRES M			



(FERR/) FERRARI E.  
Diaz-Torres M, Ferrari E;  
WPI: 2000-452412/39.  
P-PDSB: AAY97040, AAY97041, AAY97042, AAY97043, AAY97044.  
Production of proteins, such as hormones, enzymes, growth factors or cytokines, in gram-positive microorganisms containing a mutation in at least one of the genes of the opp operon gene cluster  
Disclosure; Fig 1A-M: 32pp; English.

The opp operon of Bacillus (also known as spoOK operon) encodes an oligopeptide permease that is required for the initiation of sporulation and the development of genetic competence. The opp operon is a member of the family of ATP-binding cassette (ABC) transporters involved in the import or export of oligopeptides from 3-5 amino acids. Bacillus strains containing a mutation in the opp operon produce more recombinant protein than the wild-type strain. Therefore, gram-positive microorganisms, especially Bacillus strains, containing a mutation in at least one of the genes of the opp operon can be used for heterologous protein production, especially hormones, enzymes (preferably proteases such as subtilisin), growth factors or cytokines.

Sequence 6300 BP; 1919 A; 1222 C; 1462 G; 1697 T; 0 other:

Query Match            3.5%; Score 63.6; DB 21; Length 6300;  
Best Local Similarity 43.4%; Pred. No. 2.8e-07;  
Matches 575; Conservative 0; Mismatches 724; Indels 27; Gaps 5;

QY 410 tggagtaatgaccccttaaactgaagtgaagactttatagaactcttgaaacaagtagct 469  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 804 tggctaattggagaccctgtaacgcacagatttgaatatgcttggaaatggcgacct 863  
QY 470 actcaagaagtctcaggaaatcatgtctttgttcgttcgaatccaataattgatcgaag 529  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 864 gacctaatatgaattcaatactgcgttacagctctactacataaaaggctggaagcg 923  
QY 530 attcaagaggcacacctctccatgacacattttggagtgcactctccaatgaatctaca 589  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 924 gcgaataccggaaaaggcagctagacgtgtggcagtaaaagctgtgaatgacaaaaag 983  
QY 590 ctgttttaccctgggaatccccaacctgcgattctttaaacctttgactcttcacgtc 649  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 984 ctgaagttgaattaaataaacocgactccatttcaactgaattaactgcgttctatacg 1043  
QY 650 ttcttccccttcataa-----atctcaagaagaccctgcaatcccaat----- 692  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1044 tatatgccatcataagaanaatgcagagaaaaataaaadgtgaatacaaatgcgcga 1103  
QY 693 -ctctaactatgaagcggagcttttatctctaaataatcaacaaacaaatggata 751  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1104 gatgattatgataaacgggcgttcaaatgacggcatggaacacacagcggtctatt 1163  
QY 752 aaactctcaaaaacccctcaactactataatcaaatgagtcaggtgaaactaaagattacg 811  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1164 actctgaaaaaaatgaccaglatgtggataagacaaagtcaactgaagaaaatcgat 1223  
QY 812 attcaacttactccgatgcaaacacagcgcaaaactttaatcagggaataactcaat 871  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1224 atggttatgatacaataaacaatcaggaaactaaaaaattccaagctggcgaaacttgt 1283  
QY 872 tggcaaggaacctcttggggagacgcattctccaagaaacccatctcaatttacagct 931  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1284 tggcgcggtatgcgctcggagcagc---ttccgacagaatccctgcgcgaccccgaaaaa 1340  
QY 932 aagvgggcaotcaactcttttgatgctcgaggaacccatgactcaactccaataatcaat 991  
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 1341 gacggttttcaatgttgacccatgtgcaggagtgattgttgtaaatccaactgaatgaa 1400  
QY 992 aaattccccctcaacaatatgaagcttagagaagccttagcatcagctctagataagga 1051

QY 398 aaatcagcttttggagtaatggcgaccccttaacgctgaagaccttatagaatttg 457  
||| || ||||||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Matches 325; Conservative 0; Mismatches 399; Indels 3; Gaps 2;  
Identity 44.7%; Pident NO: 2.9e-07;

Db 316 aaagatgcacaaatggagtaacggcgagccaaatcacagcaaatgatgttgaataactcttgg 375  
 Qy 458 aaacaagtagctactcaagaagctcaggaatctatgcttttgcttggaatcaataaa 517  
 Db 376 aagcgacagtgagcccaaaacaaagcttcccgagcgtatctactttgaagggttaaaa 435  
 Qy 518 aatgtacgaagatccaaaggggacacccctctccatagaccatttttgagtgcaactctct 577  
 Db 436 aattatcgtgctatgttgacgtagcaaatctaaagaaggttaggggtacagccatt 495  
 Qy 578 aatgaatcacactgtgtttacccctggaatcccaacccgctgcattcttaaaactttta 637  
 Db 496 gatgaccatacttgggaagtagcgaagctatcctatgatttttaaacattattg 555  
 Qy 638 gctcttcacagctttttcccccgttcaataatctcaagaacccctgcatacctaactctta 697  
 Db 556 ggggtaccagctttttatcttaaatgagcatttgcgaaaaacggcgcaaaaactat 615  
 Qy 698 cctatagaagcggagctttcttcttaaaataatacaacaaacaaatggataaactc 757  
 Db 616 ggtacatcagctgagtcacacacttcaaatggccttcacattagaaggttgggatggc 675  
 Qy 758 tcaaaaaacccctactactataatcaagtcagggtggaaactaaacagattacagttcac 817  
 Db 676 acgaataacttggtcctatgtagaataaaaattttgggatacgaaggaatttgcg 735  
 Qy 818 ttcattcccgatgcaaacacagcagcaaaaactatttaacagggaaactcaattggca 876  
 Db 736 ctagataaagtgaggtcccaagtagttaagaagtcacatactgggaaaaacttttcgaa 795  
 Qy 877 aggaacctctgg--ggagaacgcattctcgaagaacccctatccaaatttcaagcttaag 934  
 Db 796 ggggaagaattagatgtgttaaaatttctggagaatttggcacaagaacaggaat 855  
 Qy 935 gggcacttacactcttttgctgcaggaacccctcctatgctccacttcaatcaataaaa 994  
 Db 856 gcagcttgaaattctgaaattctcgaacgtattatcccaattaaatcaacgcaaaa 915  
 Qy 995 ttcccctcaacatatgaagcttagaagccttagcagccttagcagcttagaaggaagct 1054  
 Db 916 gatcttttggcaataagaatgcagctgcagcaatagcattatcattgaattctgagcgt 975  
 Qy 1055 ctgtctcaactatattctttagccgtgcacaaaactgcgcatcctcctacactcaaat 1114  
 Db 976 ttagtcaaaatgttttaaatgatggtctcaaaaaagcaacttggctctgtgccaacagt 1035  
 Qy 1115 attcata 1121  
 Db 1036 ttacta 1042

## RESULT 14

AAI13092  
 ID AAX13092 standard; DNA; 9021 BP.

XX  
 AC AAX13092;

XX  
 DT 19-MAR-1999 (first entry)

XX  
 DE Enterococcus faecalis genome contig SEQ ID NO:155.

XX  
 KW Enterococcus faecalis; contig; detection; Enterococcal infection;

XX  
 KW vaccine; attenuation; computer readable medium; ds.

XX  
 OS Enterococcus faecalis.

XX  
 PN WO9805055-A2.

XX  
 PD 12-NOV-1998.

XX  
 FF 04-MAY-1998; 98WO-0508985.

XX  
 PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.  
 PR 16-MAY-1997; 97US-0046655.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Barash SC, Dillon PJ, Kunsch CA;  
 PI WPI; 1999-045171/04.  
 DR  
 XX New isolated Enterococcus faecalis polynucleotides and polypeptides  
 PT - used to develop products for the detection of Enterococcus and for  
 PT use in vaccines for prevention or attenuation of Enterococcus  
 PT infection.  
 XX  
 PS Claim 1; Page 880-885; 2084pp; English.  
 XX  
 CC A computer readable medium has been developed which has recorded on it  
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
 CC AAX12938 to AAX13919 represent these nucleotide sequences which are  
 CC primary nucleotide sequences, also known as contigs. The computer-based  
 CC system can identify fragments of the Enterococcus faecalis genome with  
 CC commercial importance. The products can be used to detect the presence  
 CC of Enterococcus faecalis in samples. They can also be used for  
 CC diagnosing Enterococcal infection in an animal and monitoring  
 CC progression of disease, and for identifying agents which can be used to  
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or  
 CC another related organism, in vivo or in vitro. In particular the  
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
 CC can be used in vaccines to prevent or attenuate an Enterococcal  
 CC infection.  
 XX  
 SQ Sequence 9021 BP; 2967 A; 1582 C; 1878 G; 2581 T; 13 other;

Query Match 3.5%; Score 62.6; DB 20; Length 9021;  
 Best Local Similarity 44.7%; Pred. No. 6.3e-07;  
 Matches 325; Conservative 0; Mismatches 399; Indels 3; Gaps 2;

Qy 398 aaatcagcttttttgagtaagtcgaccccttaacagctgaagcatttatagaactctgg 457  
 Db 3376 aaagatgcaaaatgagtaacggcgagcccaatcacagaagaatgatttgaactcttgg 3435  
 Qy 458 aaacaagtagtaactcaagaagtcctcaggaatctatgttttgccttgaatccaataaa 517  
 Db 3436 aagcgacagtgagcccaaaacacagcttcccgcaagcgtattacttgaagggttaaaa 3495  
 Qy 518 aatgtacgaagatccaaagggaggaacccctcctcctagacacatttggagtgcaactctct 577  
 Db 3496 aattatcgtgctatgttgacggttagcaaatctaaagaagagtgagggttaacagccatt 3555  
 Qy 578 aatgaatctacactgtttgttacccctggaaaccccaactcgcatttcttaaacatttta 637  
 Db 3556 gatgaccataccttgggaagtagagcagctatcctctatgagttatttcaacattattg 3615  
 Qy 638 gotcttcagcttttttcccccgttcaataatctcaagaacccctgcatacctaactctta 697  
 Db 3616 ggggtaccagctttttatcctttaaataagcatttgcgaaacacggtgcaaaaactat 3675  
 Qy 698 cctatagaagcggagctttctctcctcaaaaatacaacaaaaacaaatggataaactc 757  
 Db 3676 ggtacatacagctgagtcacacacttcaatggcgttccacattagaaggttggatggc 3735  
 Qy 758 tcaaaaaacccctactactataatcaagtcagggtggaaactaaacagattacagttcac 817  
 Db 3736 acgaataacttggctcctatgtagaataaaaattttgggatacagcgaatttctgcg 3795  
 Qy 818 ttcatctccgatgcaaacacagcagcaaaaactatttaacagggaaaaactcattggca 876  
 Db 3796 ctagataaagtgaggtcccaagtagttaagaagtcacatactgggaaaaacttttcgaa 3855  
 Qy 877 aggaacctctctgg--ggagaacgcattctcgaagaacccctatccaaatttcaagcttaag 934  
 Db 3856 ggggaagaattagatgtgttaaaatttcttggaagaatttggcacaagaacaggaat 3915

QY 935 gggcaacttacactcttcttgatgcgcaggaacotcctacgtcgcacotcctcaatacaataa 994  
 Db 3916 gcagcttggaaattcgtgaaattcctggaacgtattatataccaattcaacacgcaaaa 3975  
 QY 995 tccccctcaacaatatgagcgttagagagccttagcatcagccttagataagaagct 1054  
 Db 3976 gatcttttggcaataaagaatgcacgtcgagcaatagcattatcatattcgtgagcgt 4035  
 QY 1055 ctgtctcaactatattcttaggcgcgtgcacaaactgcgacatcctcctaccataa 1114  
 Db 4036 ttacgtaaaatgttttaaatgatggctcaaaaaagcacttggtctgtgccaacaggt 4095  
 QY 1115 attcata 1121  
 Db 4096 ttacta 4102

RESULT 15  
 AAX13018  
 ID AAX13018 standard; DNA; 15747 BP.  
 XX  
 AC AAX13018;  
 XX  
 DT 19-MAR-1999 (first entry)  
 XX  
 DE Enterococcus faecalis genome contig SEQ ID NO:81.  
 XX  
 KW Enterococcus faecalis; contig; detection; Enterococcal infection;  
 KW vaccine; attenuation; computer readable medium; ds.  
 OS Enterococcus faecalis.  
 XX  
 PN W09850555-A2.  
 XX  
 PD 12-NOV-1998.  
 XX  
 PF 04-MAY-1998; 98WO-US08985.  
 XX  
 PR 14-NOV-1997; 97US-0066009.  
 PR 06-MAY-1997; 97US-0044031.  
 PR 16-MAY-1997; 97US-0046655.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Barash SC, Dillon PJ, Kunsch CA;  
 XX  
 DR WPI; 1999-045171/04.  
 XX

New Isolated Enterococcus faecalis polynucleotides and polypeptides  
 - used to develop products for the detection of Enterococcus and for  
 use in vaccines for prevention or attenuation of Enterococcus  
 infection.  
 Claim 1; Page 562-570; 2084pp; English.  
 A computer readable medium has been developed which has recorded on it  
 982 nucleotide sequences isolated from the Enterococcus faecalis genome.  
 AAX12938 to AAX13919 represent these nucleotide sequences which are  
 primary nucleotide sequences, also known as contigs. The computer-based  
 system can identify fragments of the Enterococcus faecalis genome with  
 commercial importance. The products can be used to detect the presence  
 of Enterococcus faecalis in samples. They can also be used for  
 diagnosing Enterococcal infection in an animal and monitoring  
 progression of disease, and for identifying agents which can be used to  
 modulate the growth or pathogenicity of Enterococcus faecalis, or  
 another related organism, in vivo or in vitro. In particular the  
 polypeptides encoded by the Enterococcus faecalis nucleotide sequences  
 can be used in vaccines to prevent or attenuate an Enterococcal  
 infection.  
 Sequence 15747 BP; 4949 A; 3274 C; 2618 G; 4899 T; 7 other;

Query Match 3.4%; Score 60.4; DB 20; Length 15747;  
 Best Local Similarity 48.7%; Pred. NO. 3.4e-06;  
 Matches 194; Conservative 0; Mismatches 201; Indels 3; Gaps 1;  
 QY 273 tcagccttgcacaatcatatgatgaggatagttcaagaaaataatcttccaggaaaata 332  
 Db 9377 ttaccactttaaataatgtctacgaaggaatctatcgtttagataaagacacaaacccg 9436  
 QY 333 tagagcctgctcttgcaagactactctcttctcgtgagcgaactcaacttatactttta 392  
 Db 9437 cgcctgctggtgcagcgaagaaagcagctgtttcagaaagcgttttagttcaagaagta 9496  
 QY 393 a---actgaaatcagcttttttgagtaatgagcccttaacagctgagactttatag 449  
 Db 9497 aattacgtgaagaatcaaaatggtctgatggcaaacccagttactgctgcagattacgttt 9556  
 QY 450 aatcttggaaaacagtagctactcaagaagctctcaggaaatctatgcttttgccttgaatc 509  
 Db 9557 acggttggcaacgaacagtggaatcctgctgacactgcttcagaataatgctacatgtttgaac 9616  
 QY 510 caattaaaaatgtacgaaaagatccaaagggagggacacacctctccatagaccattttgagtc 569  
 Db 9617 cagtataaaatgctgaaaaaatttctaaaggggaactacctaagaagaatgggcatta 9676  
 QY 570 actctcctaataatgactacactgtttgttaccctgggaatccccaactcgtgatttcttaa 629  
 Db 9677 aagcaatcaatgatcatgaattagaatcactctagaacacagcaacacacatttttgaog 9736  
 QY 630 aacttttagctcttcacagctcttttcccccgttcataaa 667  
 Db 9737 attataggctttccctcttcttcttgcgcgaagctcaa 9774

Search completed: July 26, 2002, 05:01:42  
 Job time: 7250 sec